

## (12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization International Bureau



(43) International Publication Date  
24 June 2004 (24.06.2004)

PCT

(10) International Publication Number  
WO 2004/053053 A2

(51) International Patent Classification: C12N  
 (21) International Application Number: PCT/EP2003/013964  
 (22) International Filing Date: 9 December 2003 (09.12.2003)  
 (25) Filing Language: English  
 (26) Publication Language: English  
 (30) Priority Data: 102 57 479.0 9 December 2002 (09.12.2002) DE  
 (71) Applicant (for all designated States except DE, US): F. HOFFMANN LA ROCHE AG [CH/CH]; Grenzacherstrasse 124, CH-4070 BASEL (CH).  
 (71) Applicant (for DE only): ROCHE DIAGNOSTICS GMBH [DE/DE]; Sandhofer Str. 116, 68305 Mannheim (DE).  
 (72) Inventors; and  
 (75) Inventors/Applicants (for US only): WATZELE, Manfred [DE/DE]; Ringstr. 2, 82362 Weilheim (DE). BUCHBERGER, Bernd [DE/DE]; Fendter Str. 6, 82380 Peissenberg (DE). PAULUS, Michael [DE/DE]; Am Forellenbach 3, 86199 Augsburg (DE).  
 (74) Agents: WEICKMANN & WEICKMANN et al.; Postfach 860 820, 81635 München (EP).  
 (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, LZ, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW, ARIPO patent (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

## Declarations under Rule 4.17:

— as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, LZ, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW, ARIPO patent (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

— of inventorship (Rule 4.17(iv)) for US only

## Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 2004/053053 A2

(54) Title: OPTIMISED PROTEIN SYNTHESIS

(57) Abstract: The invention relates to a method for the optimised production of proteins in an *in vitro* or *in vivo* expression system, and reagents suitable therefor.

Rec PTO 09 JUN 2005

**Optimierte Proteinsynthese****Beschreibung**

5

Die Erfindung betrifft ein Verfahren zur optimierten Herstellung von Proteinen in einem *in vitro* oder *in vivo* Expressionssystem sowie dafür geeignete Reagenzien.

10 Hannig, G. & Makrides, S.C. (1998) Tibtech Vol 16, pp 54-60, haben Strategien zur Optimierung der heterologen Proteinexpression in *E. coli* beschrieben. Ein entscheidender Faktor ist hierbei die Effizienz der Initiation der Translation. Die Verwendung bestimmter Codons spielt hierbei eine gewisse Rolle. So konnten George et al. (1985) DNA Vol 4, pp 273-281, zeigen, dass die Expression eines heterologen Genes gesteigert werden kann, wenn man in dem Bereich nach dem Startcodon solche Codons verwendet, die in *E. coli* Genen häufig benutzt werden. Besonders wichtig für die Translations-Initiation sind hauptsächlich strukturelle Elemente am 5'-Ende der mRNA. Von Makrides (1996) Microbiol. Rev. Vol 60, pp 512-538, wurden verschiedene Translations-Verstärker-Sequenzen, wie beispielsweise eine Sequenz aus dem T7-Phagen Gen10 Leader und eine U-reiche Sequenz aus der 5'-untranslatierten Region einiger mRNA's, wie beispielsweise des *E. coli* *atpE* Gens, beschrieben.

25 Bislang wurden keine universell einsetzbaren Translations-Initiationssequenzen beschrieben. Es wurden jedoch Strategien beschrieben, die das Potential der Sekundärstruktur-Bildung am 5'-Ende der mRNA reduzieren. Insbesondere wurde die Ribosomen-Bindungsstelle mit Adenin und Thymin Bausteinen angereichert. Stenstöm et al. (2001) Gene Vol 263, pp 273-284, zeigten, dass stark exprimierte *E. coli* Gene insbesondere bei dem dem Startcodon folgenden + 2 Codon einen hohen

30

- 2 -

Gehalt an Adeninen aufweisen. Allerdings gibt es auch für diese Regel viele positive und negative Ausnahmen.

Schließlich zeigten Pederson-Lane et al. (1997) Protein Expr. Purif. Vol 10, 5 pp 256-262, dass ein hoher GC-Gehalt unmittelbar nach dem Startcodon negativ für die Expression ist und mit Konversion der Purin Basen des 3., 4. und 5. Codons gegen Thymidin Basen die Expression der Thymidylatsynthase auf 25 % des Gesamtproteins gesteigert werden konnte.

10

Es wird angenommen, dass bei all diesen Maßnahmen die Zugängigkeit der 30S Ribosomenuntereinheit zu der messenger RNA eine entscheidende Rolle spielt. Besonders wichtig ist hierbei der freie Kontakt an die nach Shine und Dalgarno benannte Sequenz unmittelbar vor dem Startcodon und 15 der Kontakt zu dem Startcodon selbst. Sind hingegen diese Sequenzelemente in stabilen RNA-Sekundärstrukturen gebunden, so läuft die Initiation der Translation sehr ineffizient. Tessier et al. (1984) Nucl. Ac. Res. Vol 12, pp 7663-7675, zeigten in einer systematischen Untersuchung, dass durch eine gezielte Mutation diese Form von 20 Stämmchen und Schleifen (sogenannten Stem-Loops oder Hairpin-Loops) ähnelnden Sekundärstrukturen aufgelöst und damit die Effizienz der Translation erheblich gesteigert werden kann. Der Effekt dieser Sekundärstrukturen auf die Translation kann nach deren thermodynamischen Parametern berechnet werden. So bewirkt eine 25 Stabilisierung um 1,4 kcal/mol eine 10fache Reduzierung der Expression (Gold (1988) Ann. Rev. Biochem, Vol 57, pp 199-233) und eine Stabilisierung um 2,3 kcal/mol reduziert die Bindung des Ribosoms um eine Größenordnung (de Smit & van Duin (1994) J. Mol. Biol. Vol 244, pp 144-150).

30

Von Sprengart et al. (1996) EMBO Vol 15, pp 665-674, wurde als weiterer Translations-Verstärker die sogenannte „Downstream Box“, ein

Sequenzelement direkt nach dem Startcodon der T7-Gene mit Homologie zur ribosomalen 16S RNA beschrieben. Es wird angenommen, dass dieses Element durch eine Wechselwirkung der beiden homologen Basenpaare die Bindung der 30S Ribosomenuntereinheit verstrt. Allerdings eignet sich auch dieses Element nicht als universeller Translations-Verstrker.

Die Nachteile bekannter Verfahren liegen darin, dass fr jedes neue Gen eine Optimierung der 5'-Region der mRNA entweder in der 5'-untranslatierten Region oder in der translatierten Region durchgefrt werden muss, um die Codon-Verwendung zu optimieren, oder unerwnschte Sekundrstrukturen der mRNA mit Auswirkung auf die Shine-Dalgarno Sequenz oder das Startcodon zu vermeiden. Dies bedarf im Regelfall einer aufwndigen Analyse der RNA-Struktur mit entsprechenden Programmen (z.B. Mukund et al. (1999) Curr. Science Vol 76, pp 1486-1490, oder Jaeger et al. (1990) Meth. Enzymol. Vol 183, pp 281-306), sowie mehrerer PCR-Amplifikationen und Klonierungsschritte. Will man auf diese Weise eine groe Anzahl von Genen, beispielweise aus einer Genbank exprimieren, so ist dazu in jedem Falle eine genaue Kenntnis der Sequenz erforderlich, weshalb diese Methoden nicht auf unbekannte Gene angewendet werden knnen. Selbst bei Kenntnis der Sequenzen wre dieses Verfahren wesentlich aufwndiger als eine universell anwendbare Methode.

Ein weiterer Ansatz zur Verstrkung der Translation besteht darin, als universellen Translations-Enhancer ein Fusionsprotein mit einem stark exprimierten Gen zu bilden, an dessen C-terminales Ende das gewnschte Gen gesetzt wird. Als Beispiel fr den Erfolg dieser Strategie ist die Fusion mit dem Ubiquitin Gen zu nennen, wie sie von Butt et al. (1989) PNAS Vol 86, pp 2540-2544, durchgefrt wurde.

Auch dieser Ansatz kann jedoch nicht ohne Weiteres auf die Expression beliebiger Gene bertragen werden. Werden nmlich Fusionsproteine

verwendet, so wird am N-Terminus des Proteins eine mehr oder weniger große Fusion angefügt, die aufgrund der Größe und Eigenschaften des Fusionspartners mit der Funktion des gewünschten Proteins interferieren können. Je kleiner die Fusionsproteine oder Teile von diesen gewählt werden, desto geringer ist in vielen Fällen auch deren translationssteigernde Wirkung. Bei großen Fusionsproteinen zeigt sich ein weiterer Nachteil prokaryontischer Expressionssysteme: Es steigt gleichzeitig die Wahrscheinlichkeit für unvollständige Transkription oder Translation durch vorzeitige Termination oder interne Initialisierung. Auch die Wahrscheinlichkeit für proteolytischen Abbau ist erhöht.

Es besteht somit ein Bedürfnis, ein Verfahren zur optimierten Herstellung von Proteinen bereitzustellen, bei dem die Nachteile des Standes der Technik zumindest teilweise beseitigt sind.

15

Ein Gegenstand der Erfindung ist ein Verfahren zur Herstellung eines Proteins, umfassend die Schritte:

(a) Bereitstellen einer für das Protein kodierenden Nukleinsäuresequenz,

20 wobei 3'-seitig des Translations-Startcodons eine heterologe Nukleinsäuresequenz im korrekten Leseraster eingefügt wird, die so gewählt wird, dass in einem Abstand von 6-30 Nukleotiden 3'-seitig des Translations-Startcodons eine Stem-Loop-Struktur ausgebildet wird,

(b) Bereitstellen eines zur Expression des Proteins geeigneten Expressionssystems und

25 (c) Einbringen der Nukleinsäuresequenz gemäß (a) in das Expressionssystem gemäß (b) unter Bedingungen, dass das Protein synthetisiert wird.

30 Die erfindungsgemäße Lösung für ein universell optimiertes Expressionskonstrukt besteht darin, dass ein kleines heterologes DNA-Sequenzelement mit vorzugsweise maximal 201 Basenpaaren,

besonders bevorzugt maximal 45 Basenpaaren, unmittelbar nach dem Startcodon des zu exprimierenden Gens eingefügt wird, welches die Ausbildung stabiler Stem-Loop-Strukturen im Bereich der Shine-Dalgarno-Sequenz und des Startcodons weitgehend verhindert und dadurch zur optimierten Translations-Initiation und optimierter Proteinsynthese führt. Es wird somit ein Fusionsprotein gebildet, wobei vorzugsweise nur ein kleines Peptid mit maximal 67 Aminosäuren und besonders bevorzugt maximal 15 Aminosäuren an das gewünschte Protein angefügt wird.

10 Eine wichtige Voraussetzung für das heterologe DNA-Sequenzelement ist, dass es im korrekten Leseraster eingefügt wird, d.h. dass keine Rasterverschiebung im zu exprimierenden Gen stattfindet. Eine weitere wichtige Eigenschaft des heterologen DNA-Sequenzelements ist, dass sich in der transkribierten RNA in einem Abstand von 6-30 Basen, vorzugsweise 15 12-21 Basen, hinter dem Startcodon eine stabile Stem-Loop-Struktur ausbilden kann, wobei die Basenpaarung in der Stem-Loop-Struktur zumindest teilweise durch die eingefügte Sequenz bewirkt wird. Diese Stem-Loop-Struktur soll so beschaffen sein, dass sie nach der erfolgten Initiation der Translation durch das Ribosom wieder aufgelöst werden kann und somit nicht zu einem Abbruch der Translation führt. Diese durch 20 Einführung der heterologen Nukleinsäuresequenz in das Expressionskonstrukt entstandene Stem-Loop-Struktur kann sich bei nahezu jedem Gen in der gleichen Weise ausbilden und dadurch verhindern, dass die für die Translations-Initiation wichtigen Sequenzen vor dem Loop 25 größere Sekundärstrukturen mit der kodierenden Sequenz des Gens eingehen können. Der Bereich unmittelbar vor dieser Stem-Loop-Struktur und nach dem Startcodon ist vorzugsweise eine Sequenz ohne Sekundärstruktur und kann auch keine Sekundärstruktur mit der 5'-untranslatierten Region ausbilden. Besonders bevorzugt ist in dieser 30 Region eine GC-arme Sequenz, da bei einer derartigen Sequenz die Ausbildung stabiler Sekundärstrukturen mit Sequenzen innerhalb der translatierten Region minimiert wird.

Das heterologe Nukleinsäure-Sequenzelement kann in die Zielsequenz, z.B. in einen Plasmidvektor zur Expression von heterologen Genen mit bekannten Klonierungs- oder/und Amplifikationstechniken eingefügt werden. Möglich ist z.B. der Aufbau dieser Sequenz durch PCR-Primer zur 5 Klonierung des gewünschten Gens oder durch Primer, mit denen DNA-Expressionskonstrukte für die *in vitro* Proteinexpression hergestellt werden.

Das erfindungsgemäße Verfahren kann zur Herstellung und gegebenenfalls Gewinnung von Proteinen in *in vitro* Expressionssystemen eingesetzt 10 werden. Beispiele für geeignete *in vitro* Expressionssysteme sind prokaryontische *in vitro* Expressionssysteme, wie etwa Lysate von gram-negativen Bakterien, beispielsweise von *Escherichia coli*, oder gram-positiven Bakterien, wie beispielsweise *Bacillus subtilis*, oder eukaryontische *in vitro* Expressionssysteme, wie etwa Lysate von 15 Säugerzellen, wie beispielsweise von Kaninchen, Reticulocyten, humanen Tumorzelllinien, Hamsterzelllinien, oder anderen Wirbeltierzellen, wie beispielsweise Oozyten und Eiern von Fischen und Amphibien, sowie Insektenzelllinien, Hefezellen, Algenzellen oder Extrakte aus Pflanzenkeimen.

20 Alternativ kann die Herstellung des Proteins in einem *in vivo* Expressionssystem erfolgen, wobei eine prokaryontische Zelle, z.B. eine gram-negative prokaryontische Wirtszelle, insbesondere eine *E. coli* Zelle, oder eine gram-positive prokaryontische Zelle, insbesondere eine *Bacillus subtilis* Zelle, eine eukaryontische Wirtszelle, z.B. eine Hefezelle, eine 25 Insektenzelle oder eine Wirbeltierzelle, insbesondere eine Amphibien-, Fisch-, Vogel- oder Säugerzelle, oder ein nicht-humaner eukaryontischer Wirtsorganismus als Expressionssystem verwendet werden kann.

30 Die Einführung der heterologen Nukleinsäure-Sequenz in die für das gewünschte Protein kodierenden Nukleinsäure-Sequenz kann durch Standardmethoden der Molekularbiologie, z.B. durch Klonierung, wie etwa

Restriktionsspaltung oder/und Ligation, durch Rekombination oder/und durch Nukleinsäureamplifikation erfolgen. Die Nukleinsäure-Zielsequenz kann dabei auf einem geeigneten Vektor, z.B. einem Plasmidvektor zur Expression von heterologen Genen, oder auf einem Konstrukt für eine *in vitro* Proteinexpression vorliegen. Besonders bevorzugt ist eine Nukleinsäureamplifikation in einem oder mehreren Schritten, wobei durch Auswahl geeigneter Primer die heterologe Nukleinsäuresequenz und gegebenenfalls Expressionskontrollsequenzen, wie etwa Promotor, ribosomale Bindungsstellen und Terminatoren, an die für das gewünschte Protein kodierende Nukleinsäure-Sequenz angefügt werden. Besonders bevorzugt ist eine Zweistufen-PCR, wobei in einer ersten Stufe zumindest ein Teil der heterologen Nukleinsäure-Sequenz an eine Nukleinsäure-Zielsequenz, die für das gewünschte Protein kodiert, und in einem zweiten Schritt Expressionskontrollsequenzen angefügt werden. Eine bevorzugte Ausführungsform zur Durchführung einer Zweischritt-PCR ist in den Beispielen exemplarisch dargestellt.

Die heterologe Nukleinsäure-Sequenz, die in der Lage ist, eine Stem-Loop-Struktur 3'-seitig des Translations-Startcodons auszubilden, wird im korrekten Leseraster 3'-seitig des Translations-Startcodons, üblicherweise des ersten ATG-Codons, in die für das gewünschte Protein kodierende Nukleinsäure-Sequenz eingefügt. Vorzugsweise erfolgt die Einfügung in einem Abstand bis zu 6 Nukleotiden, besonders bevorzugt unmittelbar nach dem Translations-Startcodon. Eine Einfügung im "korrekten Leseraster" bedeutet dabei, dass keine Verschiebung des Leserahmen in der Protein-kodierenden Nukleinsäure-Sequenz erfolgt. Dies bedeutet wiederum, dass die Länge der heterologen Nukleinsäure-Sequenz in Nukleotiden ein Vielfaches von 3 ist. Bevorzugt ist die Länge im Bereich von 6-201 Nukleotiden, besonders bevorzugt im Bereich von 12-45 Nukleotiden.

30

Die Einfügung der heterologen Nukleinsäure-Sequenz in die Protein-kodierende Nukleinsäure-Sequenz erfolgt derart, dass eine Stem-Loop-

Struktur in einem geeigneten Abstand 3'-seitig des Translations-Codons ausgebildet wird. Der Abstand (zwischen dem letzten Nukleotid des Translationsstart-Codons und dem ersten Nukleotid des Stem) beträgt günstigerweise 6-30 Nukleotide, besonders bevorzugt 12-21 Nukleotide.

5 5'-seitig derjenigen Sequenzen, die für die Bildung der Stem-Loop-Struktur vorgesehen sind, enthält die heterologe Nukleinsäure-Sequenz vorzugsweise einen AT-reichen Bereich, d.h. einen Bereich mit einem AT-Gehalt > 50 %, insbesondere > 60 %.

10 Die Länge des Stems in der Stem-Loop-Struktur liegt vorzugsweise im Bereich von 4 bis 12 Nukleotiden, besonders bevorzugt von 5 bis 10 Nukleotiden. Der Stem der Stem-Loop-Struktur enthält vorzugsweise zwei vollständig zueinander komplementäre Abschnitte. Es können jedoch auch ein oder mehrere Basenfehlpaarungen vorhanden sein, sofern dadurch die

15 Stabilität nicht zu stark verringert wird. Die Basenpaarungen im Stem können AT- und GC-Basenpaarungen und Kombinationen davon sein. Ein Anteil der GC-Basenpaarungen von > 50 % ist bevorzugt. Die Länge des Loops beträgt vorzugsweise 2 bis 8 Nukleotide, sie ist jedoch nicht sonderlich kritisch. Die thermodynamische Stabilität der Stem-Loop-

20 Struktur ist zweckmäßigerweise hoch genug, dass die Ausbildung einer Sekundärstruktur im Bereich des ATG-Startcodons, der 15 5'-liegenden Nukleotide, welche die Shine-Dalgarno-Sequenz beinhalten, und mindestens der 5 3'-befindlichen Nukleotide verhindern. Andererseits sollte die thermodynamische Stabilität der Stem-Loop-Struktur nicht so hoch sein,

25 dass die Prozessierung des Ribosoms auf der mRNA behindert wird. Vorzugsweise liegt die thermodynamische Stabilität der Stem-Loop-Struktur im Bereich von -4 bis -15 kcal/mol.

30 Die zur Expression des gewünschten Proteins verwendeten Expressionskontrollsequenzen umfassen Promotoren, ribosomale Bindungsstellen, d.h. Shine-Dalgarno-Sequenzen für prokaryontische Expressionssysteme bzw. Kozak-Sequenzen für eukaryontische

Expressionssysteme, Enhancer, Terminatoren, Polyadenylierungs-Sequenzen etc. Dem Fachmann sind entsprechende Expressionskontrollsequenzen aus Standardlehrbüchern der Molekularbiologie, z.B. Sambrook et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor oder Ausubel et al. (1989) Current Protocols in Molecular Biology, John Wiley & Sons, New York, bekannt.

Weiterhin kann die heterologe Nukleinsäuresequenz auch Abschnitte enthalten, die für eine Aufreinigungsdomäne, z.B. eine Poly-His-Domäne, eine FLAG-Epitop-Domäne etc., oder/und für eine Proteinase-Erkennungsdomäne, z.B. eine IgA-Protease- oder Faktor-X-Domäne, kodieren. Durch die Aufreinigungsdomäne kann die Gewinnung des gewünschten Proteins, z.B. aus einem *in vitro* Translationsansatz oder einer Wirtszelle bzw. dem zur Kultivierung verwendeten Medium, vereinfacht werden. Durch Protease-Spaltung innerhalb der Proteinase-Erkennungsdomäne kann die heterologe Peptidsequenz von dem gewünschten Protein abgespalten werden.

Die heterologe Nukleinsäuresequenz oder/und die für das gewünschte Protein kodierende Nukleinsäuresequenz werden - um eine weitere Verbesserung der Expressionshöhe zu erreichen - günstigerweise so gewählt, dass sie zumindest teilweise eine an das jeweilige Expressionssystem angepasste Codon-Nutzung aufweisen.

Ein weiterer Gegenstand der Erfindung ist ein Reagenz zur Herstellung eines Proteins, umfassend

(a) eine zu der für das gewünschte Protein kodierenden Nukleinsäuresequenz heterologe Nukleinsäuresequenz, die im korrekten Leseraster in die Protein-kodierende Nukleinsäuresequenz eingefügt werden kann, und die in einem Abstand von 6-30 Nukleotiden 3'-seitig des Translations-Startcodons eine Stem-Loop-Struktur ausbilden kann, und

(b) ein zur Herstellung des Proteins geeignetes Expressionssystem.

Die heterologe Nukleinsäuresequenz kann in Form einer vollständigen Sequenz oder in Form von mehreren Teilesequenzen vorliegen.

5

Das erfindungsgemäße Verfahren und Reagenz sind insbesondere für die Synthese von Proteinen schwer exprimierbarer Gene sowie der Synthese von Proteinen ausgehend von Genbanken anwendbar, da hierbei die Erfolgsrate gegenüber der bei der Verwendung üblicher Expressionsvektoren gesteigert werden kann.

Weiterhin soll die vorliegende Erfindung durch die nachfolgenden Abbildungen und Beispiele näher erläutert werden.

15

Es zeigen:

**Figur 1** eine schematische Darstellung der zur Durchführung einer Zweischritt-PCR notwendigen Nukleinsäure-Sequenzelemente;

20

**Figur 2** eine schematische Darstellung von unterschiedlich langen Stem-Loop-Strukturen in zur Einfügung in GFP Expressionkonstrukte verwendeten heterologen Nukleinsäure-Sequenzen;

25

**Figur 3** eine Auswertung der Ergebnisse bei der Expression von GFP unter Verwendung der Hairpin-Loop-GFP-Konstrukte gemäß Figur 3 in einem RTS-Expressionssystem. 1  $\mu$ l jedes Ansatzes (Doppelbestimmungen) wurde elektrophoretisch über SDS-PAGE aufgetrennt und auf eine PVDF-Membran geblottet. Die Detektion erfolgte über DCP-Star und Lumi-Imager;

30

**Figur 4** eine schematische Darstellung von unterschiedlichen positionierten Stem-Loop-Strukturen in zur Einfügung von GFP-Expressionskonstrukte verwendeten heterologen Nukleinsäuresequenzen;

- 11 -

**Figur 5** die Expression von GFP unter Verwendung der in Figur 4 dargestellten heterologen Nukleinsäure-Sequenzen. Die Durchführung und Auswertung der Ansätze erfolgte wie in der Legende zur Figur 3 beschrieben;

5

**Figur 6** eine Auswertung der Ergebnisse bei der Expression des CIITA Gens (Wildtyp: Spur 1; Mutanten Spuren 2-10) unter Verwendung unterschiedlicher heterologer Nukleinsäuresequenzen mit Stem-Loop-Strukturen;

10

**Figur 7** eine Auswertung der Ergebnisse bei der Expression des CMV Capsid (1049) Gens (Wildtyp: Spur 1; Mutanten Spuren 2-10) unter Verwendung unterschiedlicher heterologer Nukleinsäuresequenzen mit Stem-Loop-Strukturen;

15

**Figur 8** eine Auswertung der Ergebnisse bei der Expression des Survivin Gens (Wildtyp: Spur 10; Mutanten Spuren 1-9) unter Verwendung unterschiedlicher heterologer Nukleinsäuresequenzen mit Stem-Loop-Strukturen;

20

**Figur 9** eine Auswertung der Ergebnisse bei der Expression des GFP Gens (Wildtyp: Spur 10; Mutanten Spuren 1-9) unter Verwendung unterschiedlicher heterologer Nukleinsäuresequenzen mit Stem-Loop-Strukturen;

25

**Figur 10** eine Auswertung der Ergebnisse bei der Expression des GFP und des 1049 Gens unter Verwendung unterschiedlicher heterologer Nukleinsäuresequenzen mit und ohne Stem-Loop-Strukturen;

30

**Figur 11** eine Auswertung der Ergebnisse bei der Expression des CIITA und des Survivin Gens unter Verwendung unterschiedlicher heterologer Nukleinsäuresequenzen mit und ohne Stem-Loop-Strukturen;

- 12 -

**Figur 12** eine schematische Darstellung von zwei unterschiedlichen Stem-Loop-Strukturen in den erfindungsgemäßen heterologen Sequenzen;

5 **Figur 13** eine Auswertung der Ergebnisse mit den in **Figur 12** gezeigten Stem-Loop-Strukturen;

10 **Figur 14** eine Darstellung der *in vivo* Protein Expression von RNA-Stem-Loop-Konstrukten im Vergleich zu den Wildtyp-Genen im Western Blot. Expression von drei unabhängigen Klonen der RNA-Stem-Loop-Mutanten des CMV Kapsid Proteins 1049 (Spur 1 bis 3) und des CMV Kapsid Proteins 1049 Wildtyp (Spur 4 bis 6). Expression von unabhängigen Klonen von Survivin RNA-Stem-Loop-Mutanten (Spur 7 bis 9) und des Survivin Wildtyps (Spur 10,11).

15 **Beispiele**

**Beispiel 1: Zweischritt-PCR**

20 Mit Hilfe einer Zweischritt-PCR können zu exprimierende Gene amplifiziert und mit den entsprechenden Kontrollregionen, wie T7-Promotor, T7 Gen10-Leader (g10), Ribosomaler Bindungsstelle (RBS) und T7-Terminator, versehen werden. Im ersten Schritt wird mittels eines Paares von Primern (A, B), die jeweils auf 15 Basen Länge mit dem entsprechenden Gen komplementär sind und 15 weitere Basen enthalten, die komplementär zu einem zweiten Primerpaar (C, D) sind, das Gen amplifiziert. Das zweite Primerpaar enthält alle wichtigen regulatorischen Elemente, die somit bei einer zweiten PCR-Amplifikation an das Gen angehängt werden (siehe Figur 1).

30 Über den A-Primer können bei dieser Methode Veränderungen im 5'-Bereich des Gens eingeführt werden. Bei den Hairpin-Loop-Konstrukten wurden über diesen A-Primer Hairpin-Loops mit unterschiedlicher Länge des

- 13 -

Hairpin-Loop-Stamms und in unterschiedlicher Position hinter dem Start-Codon in die Gen-Sequenz inseriert.

Primer C (SEQ ID NO. 1)

5

T7-Promotor

5'-GAAATTAATACGACTCACTATAGGGAGACCACAACGGTTCCCTCT

g10

RBS

AGAAATAATTTGTTAACTTAAAGAAGGAGATAACC-3'

10

Komplementär zu A

Primer D (SEQ ID NO. 2)

15

T7-Terminator

5'-CAAAAAACCCCTCAAGACCCGTTAGAGGCCCAAGGGGGCCGC

CAGTGTGCTGAATTGCCTTTATTA-3'

Komplementär zu B

20

Reaktionsbedingungen

Die PCR-Reaktionen wurden üblicherweise nach folgendem Schema mit dem Expand High Fidelity Kit (Roche Applied Science) im 50 µl Maßstab durchgeführt:

25

PCR 1: Template 10ng/Ansatz; Primer A 20 pmol/Ansatz; Primer B 20 pmol/Ansatz

95 °C 5 min + 20 mal (95 °C 1 min + 55 1 min + 72 °C 1 min) + 4 °C

30 PCR 2: 2 µl PCR 1; Primer C 20 pmol/Ansatz; Primer D 20 pmol/Ansatz

95 °C 5 min + 30 mal (95 °C 1 min + 50 °C 1 min +

- 14 -

72 °C 1 min) + 72 °C 10 min + 4 °C

Beide PCR-Reaktionen wurden jeweils durch Agarose-Gel-Elektrophorese überprüft und die PCR-Produkte der PCR 2 gleichzeitig mit Hilfe eines 5 DNA-Längenstandards, welcher definierte DNA-Mengen enthält, im Lumil-Imager-System quantifiziert. Die daraus erhaltenen PCR-Produkte wurden direkt als Template in RTS-Expressions-Ansätzen eingesetzt.

**Beispiel 2: Expression mit dem RTS *in vitro* Expressionssystem**

10

Die Expressionen mit dem RTS 100 HY Kit (Firma Roche Applied Science) wurden in Batch-Ansätzen zu je 50 µl nach der dem Kit beiliegenden Anleitung durchgeführt. Dabei wurden DNA-Mengen von 0,25-1 µg pro 15 Reaktions-Ansatz eingesetzt. Um die Ergebnisse einer Versuchsreihe vergleichen zu können, wurden immer gleiche Mengen des jeweiligen Templates eingesetzt. Die Ansätze wurden bei 30 °C für 4 h inkubiert.

**Beispiel 3: Expression von Hairpin-Loop-GFP-Konstrukten**

20

Es wurde am Beispiel von GFP (Grün-Fluoreszenz-Protein) untersucht, welchen Einfluss Hairpin-Loops (Haarnadel-förmige Schleifen) in der mRNA direkt nach dem Start-ATG ausüben. Dazu wurden RNA-Sequenzen ermittelt, die Hairpin-Loops (HL) mit unterschiedlichen Stammlängen ausbilden. Je länger der Stamm des Hairpin-Loops ist, um so energetisch 25 stabiler ist diese Struktur. Bei der Erstellung der Hairpin-Loops wurde darauf geachtet, dass nur in *E. coli* Genen häufig zu findende Codons verwendet wurden. Die ermittelten Sequenzen für die verschiedenen Hairpin-Loops wurden mittels mRNA-Sekundärstrukturanalyse auf ihre Stabilität im gesamten Konstrukt hin überprüft. Für Sequenzen, die 30 genügend Stabilität aufwiesen, wurden Primer erstellt, die in der beschriebenen Zweischritt-PCR gemäß Beispiel 1 eingesetzt wurden.

### Primer A:

### ohne Hairpin-Loop (SEQ ID NO. 3)

## Komplementär zu C GFP

5'-AGGAGATATACCATGACTAGCAAAGGAGAA-3'

5

### **Stammlänge 4 bp (SEQ ID NO.4)**

Komplementär zu C HL 4 bp GEP

5'-AGGAGATATACCATGACTAATTTAGTACTAGCAAAGGAGAA-3'

10 Stammlänge 5 bp (SEQ ID NO. 5)

Komplementär zu C HL 5 bp GEP

5'-AGGAGATATACCATGACTGTTTATAACGTAACTAGCAAAGGAGAA 3'

· Stammlänge 6 bp (SEQ ID NO. 6)

15 Komplementär zu C Hl. 6 bp GER

5'-AGGAGATATACCATGACTGGTCAATTACCAAGTAACTAGCAAAGGAG  
AA-3'

### Stammlänge 7 bp (SEQ ID NO. 7)

20 Komplementär zu C HL 7 bp GFP  
5'-AGGAGATATACCATGACTGCTTACATCAAGCAGTAAGTAGCAAAG  
GAGAA-3'

**Stammlänge 8 bp (SEQ ID NO. 8)**

25 Komplementär zu C HL 8 bp GFP  
5'-AGGAGATATACCATGACTGCACGTGATCGTCAGTAAGTAGCAAAG  
GAGAA-3'

**Primer B (SEQ ID NO. 9)**

30 Komplementär zu D GEP

5'-ATT CGC CCT TT TAT TAA TGT ATG ATG ATG ATG 3'

- 16 -

Eine schematische Darstellung der mRNA-Sekundärstrukturen der Hairpin-Loop-GFP-Konstrukte ist in Figur 2 gezeigt.

RTS Expression

5 Nach Expression im RTS gemäß Beispiel 2 wurde zur Verifizierung die gebildete GFP-Menge im Fluorimeter gemessen und der Western-Blot mittels CDP-Star-Detektion und Auswertung im Lumi-Imager quantitativ analysiert. Die Ergebnisse sind in Figur 3 gezeigt.

10 Es ist eindeutig erkennbar, dass die Expressionsrate mit der Stammlänge des Hairpin-Loops variiert. Bis zu einer Stammlänge von 5 bp ist die Expressionsrate relativ konstant, um dann anschließend abzufallen. Bei einer Stammlänge von 8 bp ist nahezu keine Expression mehr nachweisbar. Diese Untersuchungen bestätigen die oben erhaltenen Ergebnisse. Man 15 kann also sagen, dass ein Hairpin-Loop mit einer Stammlänge ab 6 bp oder besser mit einer Freien Energie von -7,8 kcal/mol eine Struktur darstellt, die einen erheblichen Einfluss auf die Expression ausübt. Zu erklären ist dies damit, dass diese Struktur bei den Expressionsbedingungen stabil ist und somit das vorgelagerte Start-ATG nicht frei zugänglich ist.

20

**Beispiel 4: Ermittlung des minimalen Abstandes zum Start-ATG**

Um nun zu bestimmen, bis zu welchem Abstand ein solcher Hairpin-Loop einen Einfluss auf die Expression ausübt, wurde der Hairpin-Loop mit der 25 Stammlänge 8 bp (Energie -11,8 kcal/mol) in Schritten zu je 3 Basen vom Start-ATG weg in die GFP-Sequenz verschoben. Die Sequenzen der auf diese Weise erhaltenen A-Primer war wie folgt:

Stammlänge 8 bp, 6 Basen in die GFP-Sequenz hineinverschoben (SEQ ID

30 NO. 10):

HL 8 bp                    GFP

5'-AGG...ATGACTAGCACT...GTAAAAGGAGAAGAACCTT-3'

- 17 -

Stammlänge 8 bp, 9 Basen in die GFP-Sequenz hineinverschoben (SEQ ID NO. 11):

HL 8bp GFP  
5'-AGG...ATGACTAGCAAAACT...GTAGGAGAAGAACCTTTC-3'

5

Stammlänge 8 bp, 12 Basen in die GFP-Sequenz hineinverschoben (SEQ ID NO.12):

HL 8 bp GFP  
5'-AGG...ATGACTAGCAAAGGAAC...GTAGAAGAACCTTTCACT-3'

10

Stammlänge 8 bp, 15 Basen in die GFP-Sequenz hineinverschoben (SEQ ID NO. 13):

15 HL 8 bp GFP  
5'-AGG...ATGACTAGCAAAGGAGAAAC...GTAGAACTTTCACTGG-3'

Stammlänge 8 bp, 18 Basen in die GFP-Sequenz hineinverschoben (SEQ ID NO. 14):

20 HL 8 bp GFP  
5'-AGG...ATGACTAGCAAAGGAGAAGAAAC...GTACTTTCACTGGAG  
TT-3'

25 Stammlänge 8 bp, 21 Basen in die GFP-Sequenz hineinverschoben (SEQ ID NO. 15):

HL 8bp GFP  
5'-AGG...ATGACTAGCAAAGGAGAAGAACCTACT...GTATTCAGTGGAG  
TTGTC-3'

30

Diese DNA-Konstrukte mit den in Figur 4 gezeigten Sekundärstrukturen wurden ebenfalls über Zweischritt-PCR mit den zuvor beschriebenen

Primern B, C und D synthetisiert und direkt aus der PCR-Reaktion in Expressionsansätzen als Template eingesetzt. Dabei wurde durch Quantifizierung über ein Agarose-Gel mit DNA-Marker VII und Auswertung dieses Gels im Lumi-Imager sichergestellt, dass gleiche Mengen an 5 Template eingesetzt wurden. Die Expressionsansätze wurden über einen Western-Blot ausgewertet. Die Ergebnisse sind in Figur 5 gezeigt.

Die Expressionen lassen erkennen, dass ab einem Abstand von 9 Basen zum Start-ATG die Translation der mRNA möglich ist. Es besteht aber noch 10 ein hemmender Einfluss des Hairpin-Loops. Erst ab einem Abstand von 12 Basen verläuft die Translation wieder nahezu ungehemmt. Man kann also aus diesen Ergebnissen schließen, dass das Ribosom einen Platzbedarf von 9-11 Basen nach dem Start-ATG besitzt. Des Weiteren lässt sich aus 15 diesen Ergebnissen folgern, dass ein Hairpin-Loop, der 12 oder mehr Basen nach dem Start-ATG entfernt ist, zwar einen Einfluss auf die mRNA-Sekundärstruktur aber keinen auf die Initiation der Expression hat.

#### **Beispiel 5: Einführung von Stem-Loop-Strukturen zur Auflösung ungünstiger Sekundärstrukturen**

Bei früheren Expressionsansätzen mit dem Rapid Translation System (Roche Applied Science) wurde bei einigen Genen nur eine geringe oder gar 20 keine Expression gefunden. Als Ursache hierfür wurde oft eine ungünstige RNA-Sekundärstruktur ermittelt, bei der entweder das Startcodon, oder die 25 Shine-Dalgarno-Sequenz in einer Sekundärstruktur mit der Gensequenz involviert war, und damit in gebundener Form vorlag.

Für drei dieser Gene, Survivin, Cytomegalovirus Capsid Protein 1049 (1049) und Class II Transaktivator (CIITA) wurde eine heterologe 30 Nukleinsäuresequenz mit einem Hairpin-Loop und einer Stammlänge von 7 Basen in einem Abstand von 15 Basen nach dem Startcodon eingeführt. Direkt anschließend an den Hairpin Loop wurde das Wildtyp-Gen (siehe

unten \*) ohne das Start ATG gesetzt. Vor den Hairpin- Loop wurden jeweils AT reiche Sequenzen gesetzt, welche weniger stabile Basenpaarungen als GC reiche Sequenzen ausbilden können. Außerdem wurde darauf geachtet keine für E. coli seltenen Codons innerhalb der eingeführten Sequenzen zu verwenden.

Dadurch, dass nun einerseits ein stabiler idealer Hairpin-Loop vorgegeben wird und andererseits direkt nach dem Startcodon eine Sequenz folgt, welche nicht zu Sekundärstrukturbildung neigt, sollte unabhängig vom nachfolgenden Gen eine freie Zugängigkeit für den Initiationskomplex mit der kleinen ribosomale Untereinheit an die Shine-Dalgarno-Sequenz und das Start-ATG ermöglicht werden.

Es wurden jeweils 9 verschiedene AT-reiche Sequenzen vor den Hairpin-Loops eingesetzt und mit den Wildtyp-Genen\* verglichen. Als Kontrollgen wurde das GFP cycle 3 Protein mit den gleichen Hairpin-Loops und AT-reichen Sequenzen über die in Beispiel 1 erwähnte Zweischritt-PCR synthetisiert. Die Sequenzen der A und B Primer sind unten angegeben. In Primer 1 wurden jeweils die homologen Bereiche zu Primer C unterstrichen. Die AT-reiche Sequenz wurde in kursiv, der Hairpin-Loop in Fettdruck und die Wildtyp Gensequenz in fett und unterstrichen angegeben. In Primer B wurden die zu Primer D homologen Bereiche unterstrichen und die Bereiche mit Homologie zum Wildtyp Gen fett angegeben. Als Primer D wurde abweichend zu Beispiel 1 der folgende Primer benutzt:

25

Primer D (SEQ ID NO 16):

CAAAAAACCCCTCAAGACCCGTTAGAGGCCCAAGGGGTTGGAGTA  
GAATGTTAAGGATTAGTTTATTA

Die unterstrichene Region ist homolog zu Primer C.

30

Varianten Primer A:

1049 - 1 (SEQ ID NO. 17):

AGGAGATATACCATGAAATATACATATTCTCTGCACGTGATCGTGCAG  
**GCTAACACCGCG**

**1049 - 2 (SEQ ID NO: 18):**

5 AGGAGATATACCATGAAAACATATTATTCTCTGCACGTGATCGTGCAG  
GCTAACACCGCG

**1049 - 3 (SEQ ID NO. 19):**

AGGAGATATACCATGAAATATTCTTATACACTGCACGTGATCGTGCAG  
10 GCTAACACCGCG

**1049 - 4 (SEQ ID NO. 20):**

AGGAGATATACCATGAAAATATTATTCTACACTGCACGTGATCGTGCAG  
GCTAACACCGCG

15

**1049 - 5 (SEQ ID NO. 21):**

AGGAGATATACCATGAAATATACATATTCACTGCACGTGATCGTGCAG  
GCTAACACCGCG

20

**1049 - 6 (SEQ ID NO. 22):**

AGGAGATATACCATGAAAACATATTATTCACTGCACGTGATCGTGCAG  
GCTAACACCGCG

**1049 - 7 (SEQ ID NO. 23):**

25

AGGAGATATACCATGAAATATTCAATACACTGCACGTGATCGTGCAG  
GCTAACACCGCG

**1049 - 8 (SEQ ID NO. 24):**

30

AGGAGATATACCATGAAATATTATTCAACACTGCACGTGATCGTGCAG  
GCTAACACCGCG

**1049 - 9 (SEQ ID NO. 25):**

AGGAGATATACCATGCATCATCATCATCATCTGCACGTGATCGTGCAG  
GCTAACACCGCG

5 **1049 - 10 (Wildtyp) (SEQ ID NO. 26):**

AGGAGATATACCATGGCTAACACCGCG

**1049 - Primer B (SEQ ID NO. 27):**

AGGATTAGTTTATTAATGATGATGATGATGATGGCGCCGGGTGCGCGA

10 Unterstrichenes ist homolog zu Primer D**Varianten Primer A:****Survivin - 1 (SEQ ID NO. 28):**15 AGGAGATATACCATGAAATATACATATTCTCTGCACGTGATCGTGCAG  
GGTGCCCCGACG**Survivin - 2 (SEQ ID NO. 29):**20 AGGAGATATACCATGAAAACATATTATTCTCTGCACGTGATCGTGCAG  
GGTGCCCCGACG**Survivin - 3 (SEQ ID NO. 30):**

AGGAGATATACCATGAAATATTCTTATACACTGCACGTGATCGTGCAG  
GGTGCCCCGACG

25

**Survivin - 4 (SEQ ID NO. 31):**

AGGAGATATACCATGAAATATTATTCTACACTGCACGTGATCGTGCAG  
GGTGCCCCGACG

30

**Survivin - 5 (SEQ ID NO. 32):**

AGGAGATATACCATGAAATATACATATTCACTGCACGTGATCGTGCAG  
GGTGCCCCGACG

Survivin - 6 (SEQ ID NO. 33):

AGGAGATATACCATGAAAACATATTATTCACTGCACGTGATCGTGCAG  
GGTGCCCCGACG

5    Survivin - 7 (SEQ ID NO. 34):

AGGAGATATACCATGAAATATTCAATACACTGCACGTGATCGTGCAG  
GGTGCCCCGACG

Survivin - 8 (SEQ ID NO. 35):

10    AGGAGATATACCATGAAATATTATTCAACACTGCACGTGATCGTGCAG  
GGTGCCCCGACG

Survivin - 9 (SEQ ID NO. 36):

15    AGGAGATATACCATGCATCATCATCATCATCTGCACGTGATCGTGCAG  
GGTGCCCCGACG

Survivin - 10 (A Wildtyp) (SEQ ID NO. 37):

AGGAGATATACCATGGGTGCCCGACG

20    Survivin - Primer B (SEQ ID NO. 38):

AGGATTAGTTATTAATGATGATGATGATGATGATCCATGGCAGCCAGC

CIITA - 1 (SEQ ID NO: 39):

25    AGGAGATATACCATGAAAATACATATTCTCTGCACGTGATCGTGCAG  
GAGTTGGGGCCC

CIITA - 2 (SEQ ID NO. 40):

AGGAGATATACCATGAAAACATATTATTCTCTGCACGTGATCGTGCAG  
GAGTTGGGGCCC

**CIITA - 3 (SEQ ID NO. 41):**

AGGAGATATACCATGAAATATTCTTATACTGCACGTGATCGTCAG  
GAGTTGGGGCCC

**5 CIITA - 4 (SEQ ID NO: 42):**

AGGAGATATACCATGAAATATTATTCTACACTGCACGTGATCGTCAG  
GAGTTGGGGCCC

**CIITA - 5 (SEQ ID NO. 43):**

10 AGGAGATATACCATGAAATATACATATTCACTGCACGTGATCGTCAG  
GAGTTGGGGCCC

**CIITA - 6 (SEQ ID NO. 44):**

15 AGGAGATATACCATGAAAACATATTATTCACTGCACGTGATCGTCAG  
GAGTTGGGGCCC

**CIITA - 7 (SEQ ID NO. 45):**

AGGAGATATACCATGAAATATTCAATACACTGCACGTGATCGTCAG  
GAGTTGGGGCCC

20

**CIITA - 8 (SEQ ID NO. 46):**

AGGAGATATACCATGAAATATTATTCAACACTGCACGTGATCGTCAG  
GAGTTGGGGCCC

25

**CIITA - 9 SEQ ID NO. 47):**

AGGAGATATACCATGCATCATCATCATCTGCACGTGATCGTCAG  
GAGTTGGGGCCC

**CIITA - 10 (A Wildtyp) (SEQ ID NO: 48):**

30 AGGAGATATACCATGGAGTTGGGGCCC

**CIITA Primer B (SEQ ID NO. 49):**

**AGGATTAGTTTATTATTAATGATGATGATGATGATGAGAACCCCC**

Im Folgenden sind die Sequenzen der über PCR erzeugten Expressionskonstrukte für die Mutante 1 und den Wildtyp angegeben. Die Wildtyp-Gensequenz ist fett angegeben. Am Ende des Gens wurde mit dem B-Primer jeweils ein hexa-Histidin-Tag zur Detektion mit einem spezifischen Antikörper eingeführt (unterstrichen).

**1049 – 1 (431 bp) (SEQ ID NO. 50):**

GAAATTAAATACGACTCACTATAAGGGAGACCACAAACGGTTCCCTCTAG  
AAATAATTTGTTAACTTAAGAAGGAGATATACCATGAAATATACAT  
ATTCTCTGCACGTGATCGTCAGGCTAACACCGCGCCGGGACCCAGGG  
TGGCAAACAAGCGGGACGAAAAACACCGTCACGTCGTTAACGTCGTT  
15 TCCAGCTGCCGACCGAGATATCAGAGGCCACCCACCCGGTGTGGACA  
CCATGCTGAGCAAGTACACGCGCATGTCCAGCCTGTTAACGACAAGT  
GCGCCTTAAGCTGGACCTGTCGCGATGGTAGCCGTGTCGCGCACCC  
GGCGCCCATCATCATCATCATCATTAATAAAACTAACCTAACATTCTAC  
TCCCAACCCCTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTG

20

**1049 – 10 (WildTyp) (398 bp) (SEQ ID NO. 51):**

GAAATTAAATACGACTCACTATAAGGGAGACCACAAACGGTTCCCTCTAG  
AAATAATTTGTTAACTTAAGAAGGAGATATACCATGGCTAACACCG  
CGCCGGGACCCACGGTGGCCAACAAGCGGGACGAAAAACACCGTCAC  
25 GTCGTTAACGTCGTTGGAGCTGCCGACCGAGATATCAGAGGCCACC  
CACCCGGTGTGGCCACCATGCTGAGCAAGTACACGCGCATGTCCAGC  
CTGTTAACGACAAGTGCCTTAAGCTGGACCTGTCGCGATGGTAG  
CCGTGTCGCGACCCGGGCCCATCATCATCATCATTAATAAAACTA  
ATCCTAACATTCTACTCCAACCCCTGGGCCTCTAAACGGGTCTTG  
30 AGGGGTTTTTG

Survivin – 1 (632 bp) (SEQ ID NO. 52):

GAAATTAAATACGACTCACTATAGGGAGACCACAACGGTTCCCTCTAG  
AAATAATTTGTTAACTTAAGAAGGAGATATACCATGAAATATACAT  
ATTCTCTGCACGTGATCGTGCAGGGTCCCCGACGTTCCCCCTGCCTG  
5 GCAGCCCTTCTCAAGGACCACCGCATCTCTACATTCAAGAACTGGCC  
CTTCTGGAGGGCTGCCTGCACCCCGGAGCGGATGGCCGAGGCTGG  
CTTCATCCACTGCCCACTGAGAACGAGCCAGACTTGGCCAGTGT  
CTTCTGCTTCAAGGAGCTGGAAGGCTGGAGCCAGATGACGACCCAT  
AGAGGAACATAAAAAGCATTGTCGCCGGTTGCCTTCTTCTGTCAAG  
10 AAGCAGTTGAAGAATTAACCCTGGTGAATTTTGAAGAACTGGACAGA  
GAAAGAGCCAAGAACAAAATTGCAAAGGAAACCAACAATAAGAAGAA  
AGAATTGAGGAAACTGCGAAGAAAGTGCCTGCCATCGAGCAGCT  
GGCTGCCATGGATCATCATCATCATCATCATATAAAACTAATCCTAA  
CATTCTACTCCCAACCCCTGGGCCTCTAAACGGGTCTTGAGGGTTT  
15 TTTG

Survivin – 10 (Wildtyp) (599 bp) (SEQ ID NO. 53):

GAAATTAAATACGACTCACTATAGGGAGACCACAACGGTTCCCTCTAG  
AAATAATTTGTTAACTTAAGAAGGAGATATACCATGGGTGCCCGA  
20 CGTTGCCCTGCCTGGCAGCCCTTCTCAAGGACCACCGCATCTCTAC  
ATTCAAGAACACTGCCCTTCTGGAGGGCTGCCTGCACCCCGGAGCG  
GATGGCCGAGGCTGGCTTCATCCACTGCCCACTGAGAACGAGCCAGA  
CTTGGCCCAGTGTCTCTGCTTCAAGGAGCTGGAAGGCTGGAGCC  
AGATGACGACCCATAGAGGAACATAAAAAGCATTGTCGCCGGTTGC  
25 TTTCTTCTGTCAAGAACGAGTTGAAGAATTAACCCTGGTGAATT  
TTGAAACTGGACAGAGAAAGAGCCAAGAACAAAATTGCAAAGGAAAC  
CAACAATAAGAAGAAAGAATTGAGGAAACTGCGAAGAAAGTGC  
GTGCCATCGAGCAGCTGGCTGCCATGGATCATCATCATCATCAT  
ATAAAACTAATCCTAACATTCTACTCCCAACCCCTGGGCCTCTAAAC  
30 GGGCTTGAGGGTTTTTG

CIITA - 1 (1400 bp) (SEQ ID NO. 54):

GAAATTAATACGACTCACTATAAGGGAGACCACAACGGTTCCCTCTAG  
AAATAATTTGTTAACTTAAGAAGGAGATACCATGAAATATACAT  
ATTCTCTGCACGTGATCGTCAGGAGTTGGGGCCCTAGAAGGTGGCT  
5 ACCTGGAGCTTCTAACAGCGATGCTGACCCCCCTGTGCCCTACCACTT  
CTATGACCAGATGGACCTGGCTGGAGAAGAAGAGATTGAGCTACTC  
AGAACCCGACACAGACACCATCAACTGCGACCAAGTTCAGCAGGCTGTT  
GTGTGACATGGAAGGTGATGAAGAGACCAGGGAGGCTTATGCCAATAT  
CGCGGAACTGGACCAGTATGTCTTCCAGGACTCCCAGCTGGAGGGCCT  
10 GAGCAAGGACATTTCAAGCACATAGGACCAGATGAAGTGATCGGTGA  
GAGTATGGAGATGCCAGCAGAAGTTGGCAGAAAAGTCAGAAAAGAC  
CCTTCCCAGAGGAGCTTCCGGCAGACCTGAAGCACTGGAAGCCAGCTG  
AGCCCCCCACTGTGGTGAETGGCAGTCTCCTAGTGGGACCAGTGAGCG  
ACTGCTCCACCCCTGCCCTGCCACTGCCCTGCGCTGTTCAACCAGG  
15 AGCCAGCCTCCGGCCAGATGCCCTGGAGAAAACCGACCAGATTCCC  
ATGCCCTTCTCCAGTTCCCGTTGAGCTGCCCTGAATCTCCCTGAGGGAC  
CCATCCAGTTGTCCCCACCATCTCCACTCTGCCCATGGCTCTGGCA  
AATCTCTGAGGCTGGAACAGGGTCTCCAGTATATTCATCTACCATGGT  
GAGGTGCCAGGCCAGCCAAGTACCCCCCTCCAGTGGATTCACTGTC  
20 CACGGCCTCCAAACATCTCCAGACCGGCCAGGCTCCACCAGCCCCCTC  
GCTCCATGCCACTGACCTGCCAGCATGCCCTGAACCTGCCCTGACC  
TCCCGAGCAAACATGACAGAGCACAAGACGTCCCCCACCAATGCC  
GGCAGCTGGAGAGGTCTCCAACAAGCTCCAAAATGGCCTGAGCCGGT  
GGAGCAGTTCTACCGCTCACTGCAGGACACGTATGGTGCCGAGCCCGC  
25 AGGCCCCGGATGGCATCCTAGTGGAGGTGGATCTGGTGCAGGCCAGGCT  
GGAGAGGAGCAGCAGCAAGAGCCTGGAGCGGGAACTGCCACCCGG  
ACTGGGCAGAACGGCAGCTGGCCAAGGAGGCCTGGCTGAGGTGCTG  
TTGGCTGCCAAGGAGCACCGGCGGCCGCTGACTCGAGCGAGCTCCC  
GGGGGGGGTTCTCATCATCATCATCATTAATAATAACTAATCCTT  
30 AACATTCTACTCCAAACCCCTGGGCCTCTAAACGGGTCTTGAGGGG  
TTTTTG

CIITA - 10 WT 1367 bp (SEQ ID NO. 55):

GAAATTAAATACGACTCACTATAGGGAGACCACAACGGTTCCCTCTAG  
AAATAATTGTTAACTTAAGAAGGAGATATACCATGGAGTTGGGGC  
5 CCCTAGAAGGTGGCTACCTGGAGCTTAAACAGCGATGCTGACCCCC  
TGTGCCTCTACCACTTCTATGACCAGATGGACCTGGCTGGAGAAGAAG  
AGATTGAGCTCTACTCAGAACCCGACACAGACACCATCAACTGCGACC  
AGTTCAGCAGGCTGTTGTGACATGGAAGGTGATGAAGAGACCAAGGG  
AGGCTTATGCCAATATCGCGGAACGGACAGTGTCTCCAGGACT  
10 CCCAGCTGGAGGGCCTGAGCAAGGACATTTCAAGCACATAGGACCAAG  
ATGAAGTGATCGGTGAGAGTATGGAGATGCCAGCAGAAGTGGCAG  
AAAAGTCAGAAAAGACCCCTCCAGAGGGAGCTCCGGCAGACCTGAA  
GCACTGGAAGCCAGCTGAGCCCCACTGTGGTACTGGCAGTCTCCT  
AGTGGGACCAGTGAGCGACTGCTCCACCCCTGCCCTGCCACTGCC  
15 TCGCCTGTTCAACCAGGAGCCAGCCTCCGGCCAGATGCCCTGGAGAA  
AACCGACCAGATTCCCATGCCCTTCTCCAGTTCTCGTTGAGCTGCCTG  
AATCTCCCTGAGGGACCCATCCAGTTGTCCCCACCATCTCCACTCTGC  
CCCATGGGCTCTGGCAAATCTCTGAGGCTGGAACAGGGGTCTCCAGTA  
TATTCTACCATGGTGAGGTGCCCCAGGCCAGCCAAGTACCCCTC  
20 CCAGTGGATTCACTGTCCACGCCCTCCAACATCTCCAGACCGGCCAG  
GCTCCACCAGCCCCCTCGCTCCATCAGCCACTGACCTGCCAGCATGC  
CTGAACCTGCCCTGACCTCCCGAGCAAACATGACAGAGCACAAGACGT  
CCCCCACCCAAATGCCCGGCAGCTGGAGAGGTCTCCAACAAGCTTCAA  
AATGGCCTGAGCCGGTGGAGCAGTTCTACCGCTCACTGCAGGACACGT  
ATGGTGCCAGCCCGCAGGCCGGATGGCATCCTAGTGGAGGTGGATC  
25 TGGTGCAGGCCAGGCTGGAGAGGAGCAGCAGCAAGAGCCTGGAGCGG  
GAACTGGCCACCCCGACTGGCAGAACGGCAGCTGGCCCAAGGAGG  
CCTGGCTGAGGTGCTGGCTGCCAAGGAGCACCAGGGCCCGTCG  
ACTCGAGCGAGCTCCGGGGGGTTCTCATCATCATCATCATCAT  
ATAATAAAACTAACCTAACATTCTACTCCAAACCCCTGGGGCTCTA  
30 AACGGGTCTTGAGGGTTTTG

GFP CyC3 - 1 (938 bp) (SEQ ID NO. 56):

GAAATTAATACGACTCACTATAGGGAGACCACAACGGTTCCCTCTAG  
AAATAATTTGTTAACTTAAGAAGGAGATATACCATGAAATATACAT  
ATTCTCTGCACGTGATCGTGCAGACTAGCAAAGGAGAAGAACCTTCA  
5 CTGGAGTTGCCAACATTCTGTTGAATTAGATGGTATGTTAATGGCA  
CAAATTTCTGTCAGTGGAGAGGGTGAAGGTGATGCTACATACGGAAA  
GCTTACCCCTAAATTATTCACACTGGAAAACACCTGTTCCATGG  
CCAACACTGTCACTACTTCTCTTATGGTGTCAATGCTTTCCCGTTA  
TCCGGATCATATGAAACGGCATGACTTTCAAGAGTGCCATGCCGA  
10 AGGTTATGTACAGGAACGCACATATCTTCAAAGATGACGGAACTA  
CAAGACGCGTGCTGAAGTCAAGTTGAAGGTGATACCCCTGTTAATCG  
TATCGAGTTAAAAGGTATTGATTTAAAGAAGATGGAAACATTCTCGG  
ACACAAACTCGAGTACAACACTATAACTCACACAATGTATACATCACGGC  
AGACAAACAAAAGAATGGAATCAAAGCTAACCTCAAAATTGCCACA  
15 ACATTGAAGATGGATCCGTTCAACTAGCAGACCATTATCAACAAAATA  
CTCCAATTGGCGATGGCCCTGTCCTTACCAAGACAACCATTACCTGTC  
GACACAATCTGCCCTTCGAAAGATCCAACGAAAAGAGAGACCACAT  
GGTCCTTCTTGAGTTGTAACAGCTGCTGGATTACACATGGCATGGAT  
GAACATACAAACCCGGGGGGGTTCTCATCATCATCATCATCATTAA  
20 TAAACTAATCCTAACATTCTACTCCCAACCCCTGGGGCCTAAACG  
GGTCTTGAGGGGTTTTTG

pIVEX-GFP CyC3 - 10 905 bp (SEQ ID NO. 57):

GAAATTAATACGACTCACTATAGGGAGACCACAACGGTTCCCTCTAG  
25 AAATAATTTGTTAACTTAAGAAGGAGATATACCATGACTAGCAAA  
GGAGAAGAACCTTCACTGGAGTTGTCAGTGGAGAGGGTGAAGGTG  
GTGATGTTAATGGCACAAATTCTGTCAGTGGAGAGGGTGAAGGTG  
ATGCTACATACGGAAAGCTTACCCCTAAATTATTCACACTGGAAA  
ACTACCTGTTCCATGGCCAACACTGTCACACTTCTCTTATGGTGTTC  
30 AATGCTTTCCGTTATCGGATCATATGAAACGGCATGACTTTCAA  
GAGTGCCATGCCGAAGGTTATGTACAGGAACGCACATATCTTCAA  
AGATGACGGAACTACAAGACGCGTGCTGAAGTCAAGTTGAAGGTGA

TACCCCTGTTAACGTATCGAGTTAAAGGTATTGATTTAAAGAAGAT  
GGAAACATTCTCGGACACAAACTCGAGTACAACATAACTCACACAAT  
GTATACATCACGGCAGACAAACAAAAGAATGGAATCAAAGCTAACTTC  
AAAATTGCCACAAACATTGAAGATGGATCCGTTCAACTAGCAGACCAT  
5 TATCAACAAAATCTCCAATTGGCGATGGCCCTGTCCTTACCAAGACA  
ACCATTACCTGTCGACACAATCTGCCCTTCGAAAGATCCAACGAAA  
AGAGAGACCACATGGTCCTTCTTGAGTTGTAACAGCTGCTGGGATTAC  
ACATGGCATGGATGAACTATAACAAACCCGGGGGGTTCTCATCATCA  
TCATCATCATTAATAAAACTAACATTCTAACATTCTACTCCAACCCCTG  
10 GGGCCTCTAAACGGGTCTTGAGGGGTTTTG

Man kann an den in den Figuren 6 bis 9 gezeigten Expressionen erkennen, dass die mit den Stem-Loop-Strukturen synthetisierten DNA Templates in allen Fällen zu einer Proteinsynthese führten, während beim Wildtypgen 15 keine Proteinsynthese stattfand. Die Mutante 9 mit der Hexa-Histidin Sequenz wird dabei etwas schlechter als die übrigen AT-reichen Sequenzen exprimiert, hat aber den Vorteil, dass man das entstehende Protein über diese Markierung mit sechs Histidinresten an Ni-NTA-Chelat Säulen aufreinigen kann. Auch im Falle des GFP-Gens, einem ohnehin gut 20 exprimierten Gen, führten die Stem-Loop-Konstrukte zu einer Ausbeutesteigerung.

#### **Beispiel 6: Entfernen der Stem-Loop-Struktur zum Nachweis der Funktion**

25 Um die Wirkung der Stem-Loop-Struktur von der Wirkung der eingeführten AT-reichen Sequenz zu unterscheiden wurden von je zwei der Mutanten eine identische PCR jedoch ohne den Teil des Stem-Loops hergestellt und im direkten Vergleich mit den Stem-Loop-Mutanten exprimiert.

30 An diesen Beispielen lässt sich deutlich der Einfluss der Stem-Loop-Struktur ersehen. Während beim GFP die AT-reiche Sequenz alleine zu einer

gesteigerten Expression führt, bringt die Stem-Loop-Sequenz bei den schwer exprimierbaren Genen den entscheidenden Beitrag.

**Beispiel 7: Modifikation der Stem-Loop-Struktur zur Ermittlung der für die Funktion wichtigen Eigenschaften**

Um die Wirkung der GC-Basen innerhalb der Stem-Loop-Struktur zu ermitteln, wurde deren Sequenz durch eine AT-reiche Sequenz mit der gleichen freien Energie wie der des GC-reichen Stem-Loops ersetzt.

Für die Beispiele Survivin, CIITA und 1049 wurden dafür statt der ursprünglichen Stem-Loop-Sequenz CTG.CAC.GTG.ATC.GTG.CAG mit (G = -9,8 kcal/mol und einer Stammlänge von 7 Basenpaaren ein neuer Stem-Loop (Loop') mit der Sequenz CAG.ACA.AAT.AGA.TAT.TTG.TCT.GTA mit (G = -9,8 kcal/mol und einer Stammlänge von 9 Basenpaaren mit der AT-reichen Sequenz von Mutante 1 kombiniert. Die beiden Strukturen sind in Figur 12 gezeigt.

Man kann erkennen, dass beide Stem-Loop-Varianten die Expression gegenüber den jeweiligen Wildtypgenen deutlich steigern oder erst ermöglichen. Die GC-reichen Stem-Loop-Varianten weisen dabei eine etwas deutlichere Expressionssteigerung auf.

**Beispiel 8: *In vivo* Protein Expression**

Die PCR-Produkte aus Beispiel 5 mit dem Expressionskonstrukt für das Wildtypgen des Cytomegalovirus Capsid Protein 1049 sowie für das Survivin Wildtypgen wurden in pBAD-TOPO (Invitrogen, Carlsbad, USA) Vektoren kloniert. Ebenso wurden jeweils die Expressionskonstrukte für Mutanten des Gens für das Cytomegalovirus Capsid Protein 1049 und das Survivin Gen in diesen Vektor kloniert. Danach wurden die Plasmide in B21 pLyS (Stratagene, Amsterdam, Netherlands) Stämme transformiert und auf

auf LB Platten mit 100  $\mu$ g/ml Carbenicillin und 34  $\mu$ g/ml Chloramphenicol ausgestrichen. Die Inserts wurden durch Sequenzierung überprüft. Für die *in vivo* Protein Expression wurden je 3 Kolonien isoliert und in 4 ml Medium bei 37 °C für 5 h angezogen. Bei Erreichen einer Zelldichte von 10<sup>8</sup> 5 Zellen/ml, wurde die Expression durch Zugabe von 1mM IPTG induziert und weitere 2 h inkubiert. Je 1 ml Zellsuspension davon wurde abzentrifugiert (3 min 14000 rpm) und der Niederschlag in SDS Probenpuffer für 20 min bei 95 °C 1400 rpm auf einem Thermoschüttler erhitzt. Je 10  $\mu$ l davon 10 wurden auf ein SDS Gel aufgetragen und mit einem Western Blot wie in Beispiel 5 analysiert.

Die Expressionen in Figur 14 zeigen, dass die Stem-Loop-Konstrukte für die beiden untersuchten Gene Cytomegalovirus Capsid Protein 1049 sowie Survivin auch *in vivo* eine wesentliche höhere Expression als die 15 Wildtypgene aufweisen. Dies beweist, dass sich die Ergebnisse der *in vitro* Expression auf die *in vivo* Expression übertragen lassen.

**Ansprüche**

1. **Verfahren zur Herstellung eines Proteins, umfassend die Schritte:**
  - 5 (a) **Bereitstellen einer für das Protein kodierenden Nukleinsäuresequenz, wobei 3'-seitig des Translations-Startcodons eine heterologe Nukleinsäuresequenz im korrekten Leseraster eingefügt wird, die so gewählt wird, dass in einem Abstand von 6-30 Nukleotiden 3'-seitig des Translations-Startcodons eine Stem-Loop-Struktur ausgebildet wird,**
  - (b) **Bereitstellen eines zur Expression des Proteins geeigneten Expressionssystems und**
  - 10 (c) **Einbringen der Nukleinsäuresequenz gemäß (a) in das Expressionssystem gemäß (b) unter Bedingungen, dass das Protein synthetisiert wird.**
2. **Verfahren nach Anspruch 1, weiterhin umfassend das Gewinnen des Proteins.**
- 20 3. **Verfahren nach Anspruch 1 oder 2, dadurch gekennzeichnet, dass die eingefügte heterologe Nukleinsäuresequenz eine Länge bis zu 201 Nukleotiden aufweist.**
- 25 4. **Verfahren nach Anspruch 3, dadurch gekennzeichnet, dass die eingefügte heterologe Nukleinsäuresequenz eine Länge bis zu 45 Nukleotiden aufweist.**

- 33 -

5. Verfahren nach einem der Ansprüche 1 bis 4,  
**dadurch gekennzeichnet,**  
dass die Stem-Loop-Struktur in einem Abstand von 12-21  
Nukleotiden 3'-seitig des Startcodons ausgebildet wird.

5

6. Verfahren nach einem der Ansprüche 1 bis 5,  
**dadurch gekennzeichnet,**  
dass die Länge des Stem in der Stem-Loop-Struktur im Bereich von  
4-12 Nukleotiden ist.

10

7. Verfahren nach einem der Ansprüche 1 bis 6,  
**dadurch gekennzeichnet,**  
dass der Bereich der heterologen Nukleinsäuresequenz, der 5'-seitig  
der Stem-Loop-Struktur liegt, selbst keine Sekundärstruktur ausbildet  
15 und keine Sekundärstruktur mit der 5'-untranslatierten Region der für  
das herzustellende Protein kodierenden Nukleinsäuresequenz  
eingehen kann.

15

8. Verfahren nach einem der Ansprüche 1 bis 7,  
**dadurch gekennzeichnet,**  
dass der Bereich der heterologen Nukleinsäuresequenz, der 5'-seitig  
der Stem-Loop-Struktur und 3'-seitig des ATG-Startcodons liegt,  
einen GC-Gehalt von < 50 % aufweist.

20

25 9. Verfahren nach einem der Ansprüche 1 bis 8,  
**dadurch gekennzeichnet,**  
dass man ein *in vitro* Expressionssystem verwendet.

25

10. Verfahren nach Anspruch 9,  
**dadurch gekennzeichnet,**  
dass man ein prokaryontisches *in vitro* Expressionssystem  
verwendet.

30

11. Verfahren nach Anspruch 10,  
**dadurch gekennzeichnet,**  
dass das prokaryontische *in vitro* Expressionssystem Lysate von  
gram-negativen Bakterien, insbesondere von *Escherichia coli*, oder  
5 gram-positiven Bakterien, insbesondere *Bacillus subtilis*, enthalten.

12. Verfahren nach Anspruch 9,  
**dadurch gekennzeichnet,**  
dass man ein eukaryontisches *in vitro* Expressionssystem  
10 verwendet.

13. Verfahren nach Anspruch 12,  
**dadurch gekennzeichnet,**  
dass das eukaryontische *in vitro* Expressionssystem Lysate von  
15 Säugerzellen, insbesondere von Kaninchen, Reticulocyten, humanen  
Tumorzelllinien, Hamsterzelllinien, oder anderen Wirbeltierzellen,  
insbesondere Oozyten und Eiern von Fischen und Amphibien, sowie  
Insektenzelllinien, Hefezellen, Algenzellen oder Extrakte aus  
Pflanzenkeimen enthält.

20 14. Verfahren nach einem der Ansprüche 1 bis 8,  
**dadurch gekennzeichnet,**  
dass man ein prokaryontisches *in vivo* Expressionssystem  
verwendet.

25 15. Verfahren nach Anspruch 14,  
**dadurch gekennzeichnet,**  
dass man eine prokaryontische Wirtszelle als Expressionssystem  
verwendet.

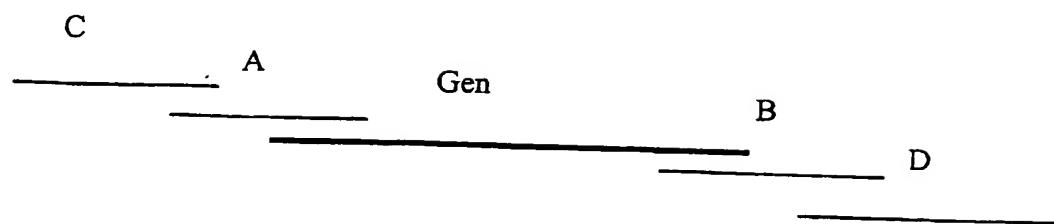
16. Verfahren nach Anspruch 15,  
**dadurch gekennzeichnet,**  
dass man eine gram-negative prokaryontische Wirtszelle,  
insbesondere eine E.coli Zelle, oder eine gram-positive  
5 prokaryontische Wirtszelle, insbesondere eine *Bacillus subtilis* Zelle,  
verwendet.
17. Verfahren nach Anspruch 14,  
**dadurch gekennzeichnet,**  
10 dass man eine eukaryontische Wirtszelle als Expressionssystem  
verwendet.
18. Verfahren nach Anspruch 17,  
**dadurch gekennzeichnet,**  
15 dass man eine Hefezelle, eine Insektenzelle oder eine Wirbeltierzelle,  
insbesondere eine Amphibien-, Fische-, Vogel- oder Säugerzelle,  
verwendet.
19. Verfahren nach Anspruch 14,  
**dadurch gekennzeichnet,**  
20 dass man einen nicht-humanen eukaryontischen Wirtsorganismus als  
Expressionssystem verwendet.
20. Verfahren nach einem der Ansprüche 1 bis 19,  
**dadurch gekennzeichnet,**  
25 dass das Bereitstellen der für das Protein kodierenden  
Nukleinsäuresequenz durch Klonierung, Rekombination oder/und  
Amplifikation erfolgt.
- 30 21. Verfahren nach Anspruch 20,  
**dadurch gekennzeichnet,**  
dass das Bereitstellen eine Zweistufen-PCR umfasst.

22. Verfahren nach einem der Ansprüche 1 bis 21,  
dadurch gekennzeichnet,  
dass die für das herzustellende Protein kodierende  
5 Nukleinsäuresequenz oder/und die heterologe Nukleinsäuresequenz  
zumindest teilweise eine an das jeweiligen Expressionssystem  
angepasste Codon-Nutzung aufweisen.

23. Verfahren nach einem der Ansprüche 1 bis 22,  
dadurch gekennzeichnet,  
10 dass die heterologe Nukleinsäuresequenz einen für eine  
Aufreinigungsdomäne oder/und einen für eine Proteinase-  
Erkennungsdomäne kodierenden Abschnitt enthält.

24. Reagenz zur Herstellung eines Proteins, umfassend  
15 (a) eine zu der für das Protein kodierenden Nukleinsäuresequenz  
heterologe Nukleinsäuresequenz, die im korrekten Leseraster  
in die Protein-kodierende Nukleinsäuresequenz eingefügt  
werden kann, und die in einem Abstand von 6-30 Nukleotiden  
3'-seitig des Translations-Startcodons eine Stem-Loop-  
20 Struktur ausbilden kann, und  
(b) ein zur Herstellung des Proteins geeignetes  
Expressionssystem.

Figur 1



**Figur 2**

HL mit Stammlänge 6bp: -7,8 kcal/mol

A  
A U  
C U  
U-A  
G-C  
G-C  
U-A  
C-G  
A-U

5'-....AGGAGAUAUACCAUG ACUAGCAAAGGAGAA... -3'

HL mit Stammlänge 7bp: -8,6 kcal/mol

C A  
A U  
U C  
U-A  
U-A  
C-G  
G-C  
U-A  
C-G  
A-U

5'-....AGGAGAUAUACCAUG ACUAGCAAAGGAGAA... -3'

**Figur 2 (Fortsetzung)**

HL mit Stammlänge 4bp: -4,1 kcal/mol

U U  
A U  
A-U  
U-A  
C-G  
A-U

5'-.....AGGAGAUAUACCAUG ACUAGCAAAGGAGAA... -3'

HL mit Stammlänge 5bp: -4,4 kcal/mol

U A  
U U  
U-A  
G-C  
U-A  
C-G  
A-U

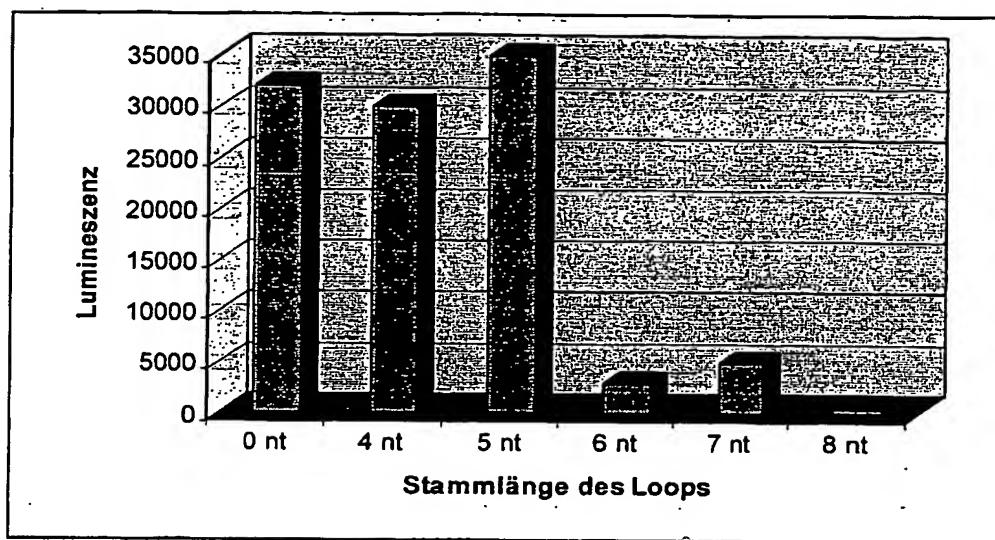
5'-.....AGGAGAUAUACCAUG ACUAGCAAAGGAGAA... -3'

HL mit Stammlänge 8bp: -14,8 kcal/mol

G A  
U U  
G-C  
C-G  
A-U  
C-G  
G-C  
U-A  
C-G  
A-U

5'-.....AGGAGAUAUACCAUG ACUAGCAAAGGAGAA... -3'

Figur 3



**Figur 4**

HL mit Stammlänge 8bp, 6 Basen nach Start-ATG:

G A  
U    U  
G-C  
C-G  
A-U  
C-G  
G-C  
U-A  
C-G  
A-U

5'-.....AGGAGAUAUACCAUGACUAGC    AAAGGAGAA... -3'

HL mit Stammlänge 8bp, 9 Basen nach Start-ATG:

G A  
U    U  
G-C  
C-G  
A-U  
C-G  
G-C  
U-A  
C-G  
A-U

5'-.....AGGAGAUAUACCAUGACUAGCAA    GGAGAA... -3'

HL mit Stammlänge 8bp, 12 Basen nach Start - ATG:

G A  
U    U  
G-C  
C-G  
A-U  
C-G  
G-C  
U-A  
C-G  
A-U

**Figur 4 (Fortsetzung)**

5'-.....AGGAGAUAUACCAUGACUAGCAAAGGA    GAA... -3'

                  G A  
                  U    U  
                  G-C  
                  C-G  
                  A-U  
                  C-G  
                  G-C  
                  U-A  
                  C-G  
                  A-U

5'-.....AGGAGAUAUACCAUGACUAGCAA A    GGAGAA... -3'

HL mit Stammlänge 8bp, 12 Basen nach Start - ATG:

                  G A  
                  U    U  
                  G-C  
                  C-G  
                  A-U  
                  C-G  
                  G-C  
                  U-A  
                  C-G  
                  A-U

5'-.....AGGAGAUAUACCAUGACUAGCAAAGGA    GAA... -3'

7/16

**Figur 4 (Fortsetzung)**

- HL mit Stammlänge 8bp, 15 Basen nach Start-ATG:

|     |   |
|-----|---|
| G   | A |
| U   | U |
| G-C |   |
| C-G |   |
| A-U |   |
| C-G |   |
| G-C |   |
| U-A |   |
| C-G |   |
| A-U |   |

5'-.....AGGAGAUAUACCAUGACUAGCAAAGGAGAA      GAA... -3'

- HL mit Stammlänge 8bp, 18 Basen nach Start-ATG:

|     |   |
|-----|---|
| G   | A |
| U   | U |
| G-C |   |
| C-G |   |
| A-U |   |
| C-G |   |
| G-C |   |
| U-A |   |
| C-G |   |
| A-U |   |

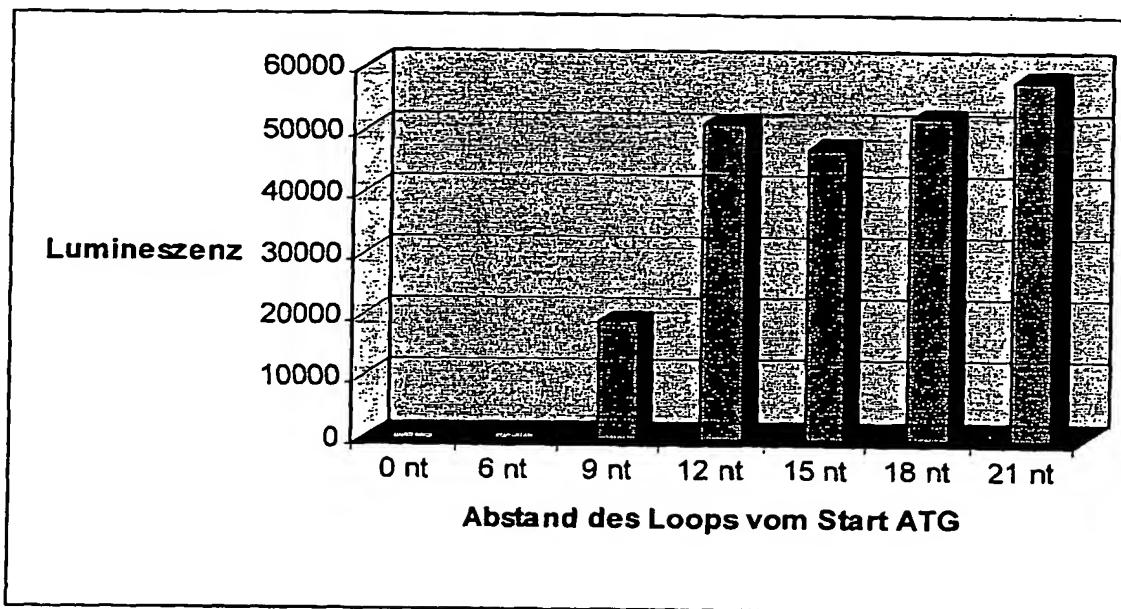
5'-.....AGGAGAUAUACCAUGACUAGCAAAGGAGAAAGAA      CTT... -3'

- HL mit Stammlänge 8bp, 21 Basen nach Start-ATG:

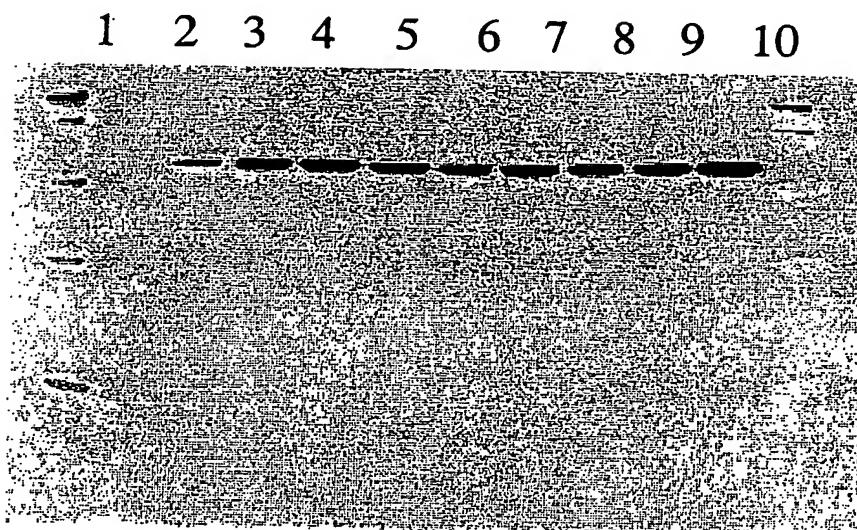
|     |   |
|-----|---|
| G   | A |
| U   | U |
| G-C |   |
| C-G |   |
| A-U |   |
| C-G |   |
| G-C |   |
| U-A |   |
| C-G |   |
| A-U |   |

5'-.....AGGAGAUAUACCAUGACUAGCAAAGGAGAAACTT      TTC... -3'

Figur 5

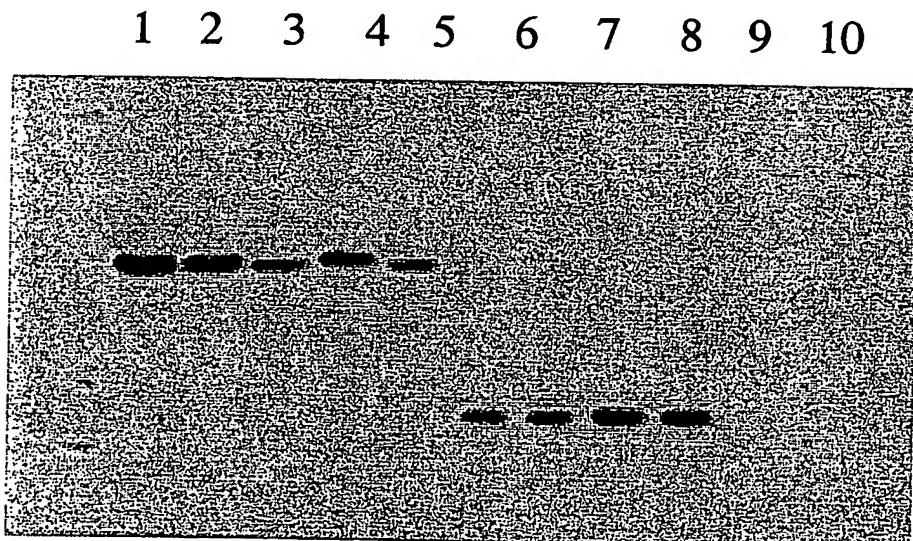


9/16

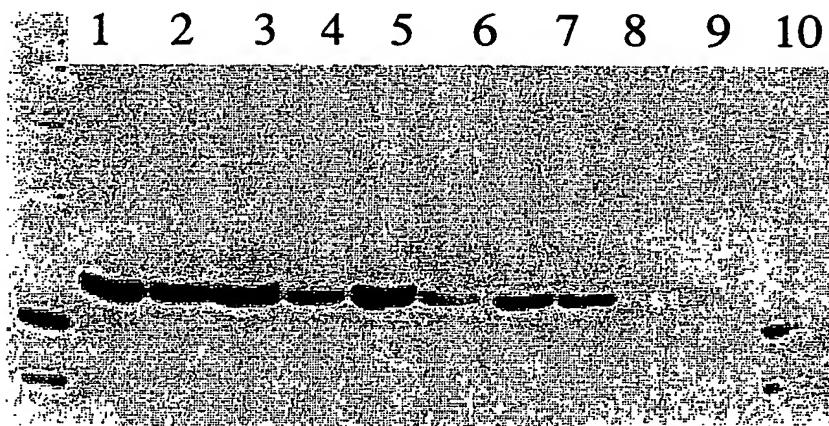
**Figur 6****Spur 1-9= Mutante 9-1 Spur 1 entspricht der Wildtypsequenz**

**Figur 7****GFP/1049**

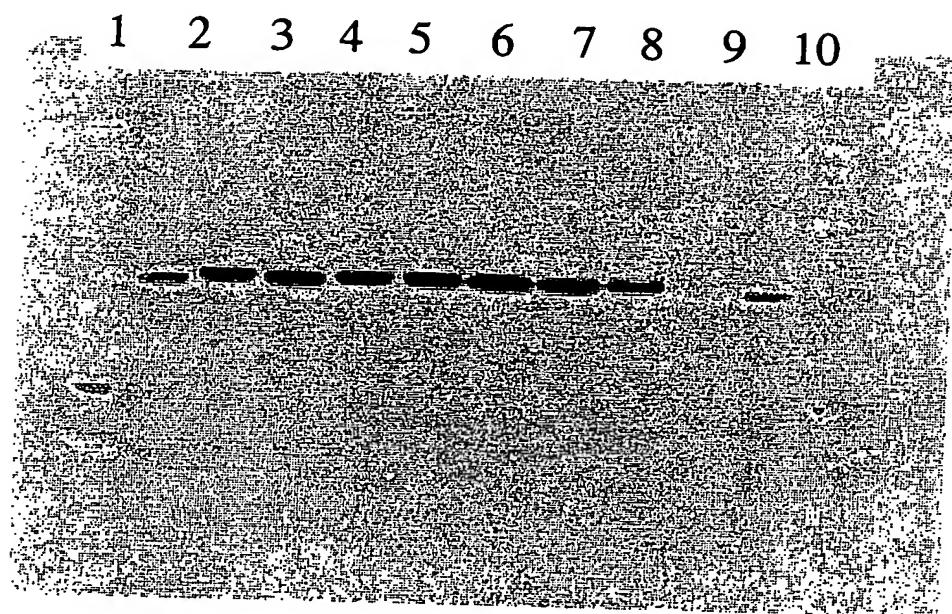
**Spur 1,2= Mutante 1,2 GFP ohne Loop. Spur 3,4= Mutante 1,2 GFP mit Loop.**  
**Spur 5 entspricht der GFP Wildtypsequenz. Spur 6,7= Mutante 1,2 1049**  
**ohne Loop. Spur 8,9= Mutante 1,2 1049 mit Loop. Spur 10 entspricht der**  
**1049 Wildtypsequenz.**



11/16

**Figur 8****Survivin****Spur 1-9 = Mutante 1-9 Spur 10 entspricht der Wildtypsequenz**

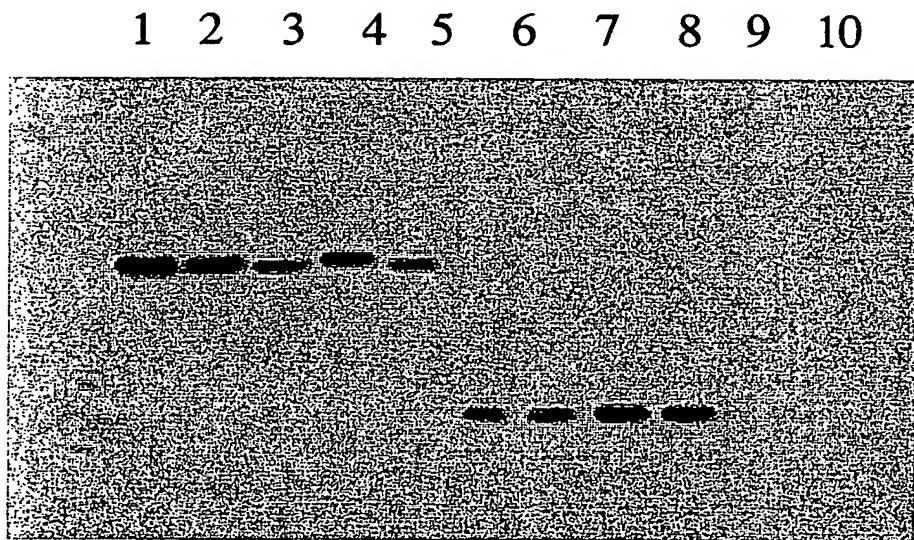
12/16

**Figur 9****GFP****Spur 1-9 = Mutante 1-9 Spur 10 entspricht der Wildtypsequenz**

13/16

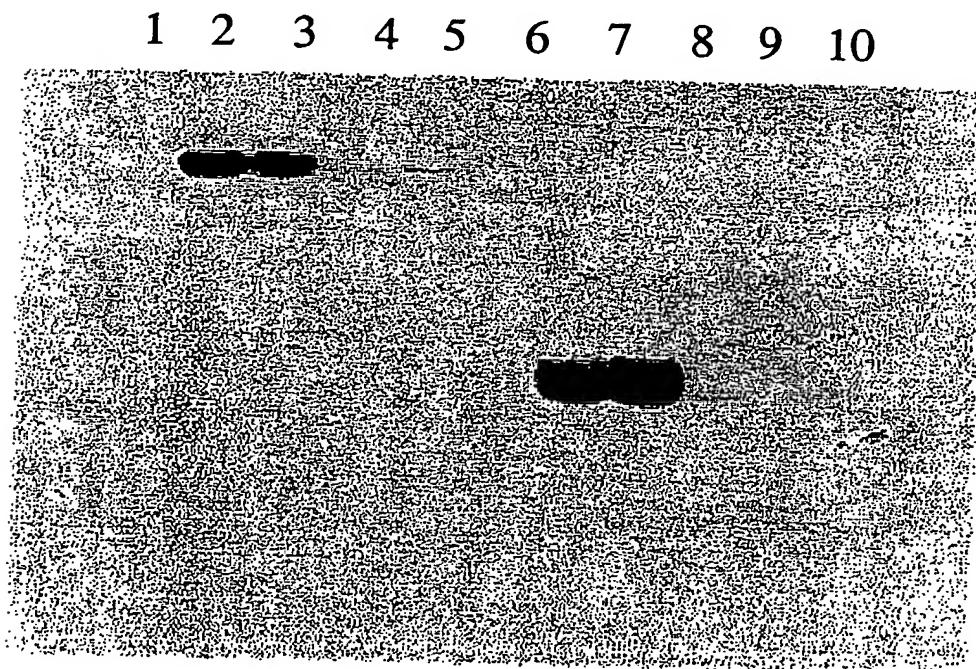
**Figur 10****GFP/1049**

**Spur 1,2= Mutante 1,2 GFP ohne Loop. Spur 3,4= Mutante 1,2 GFP mit Loop.  
Spur 5 entspricht der GFP Wildtypsequenz. Spur 6,7= Mutante 1,2 1049  
ohne Loop. Spur 8,9= Mutante 1,2 1049 mit Loop. Spur 10 entspricht der  
1049 Wildtypsequenz.**



**Figur 11****CIITA/Survivin**

**Spur 1 entspricht der CIITA Wildtypsequenz. Spur 2,3= Mutante 1,2 CIITA mit Loop. Spur 4,5= Mutante 1,2 CIITA ohne Loop. Spur 6 entspricht der Survivin Wildtypsequenz. Spur 7,8= Mutante 1,2 Survivin mit Loop. Spur 9,10= Mutante 1,2 Survivin ohne Loop..**



Figur 12

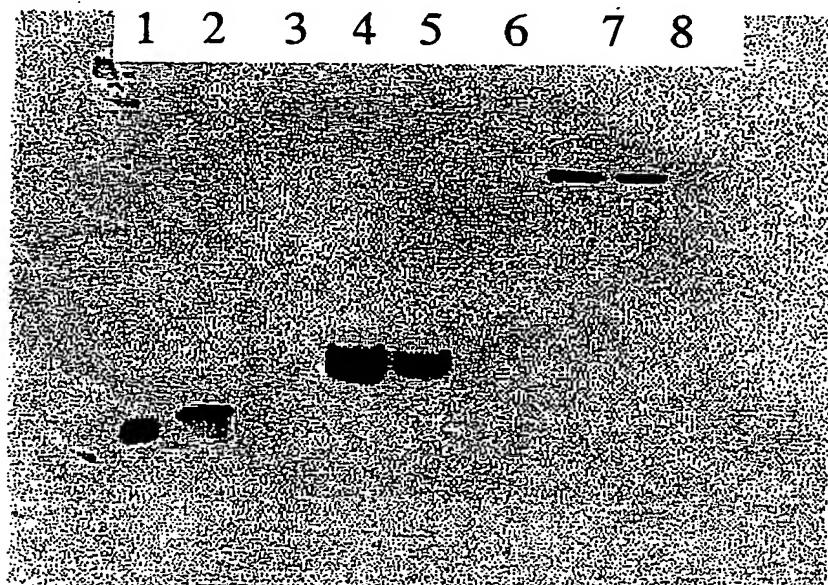
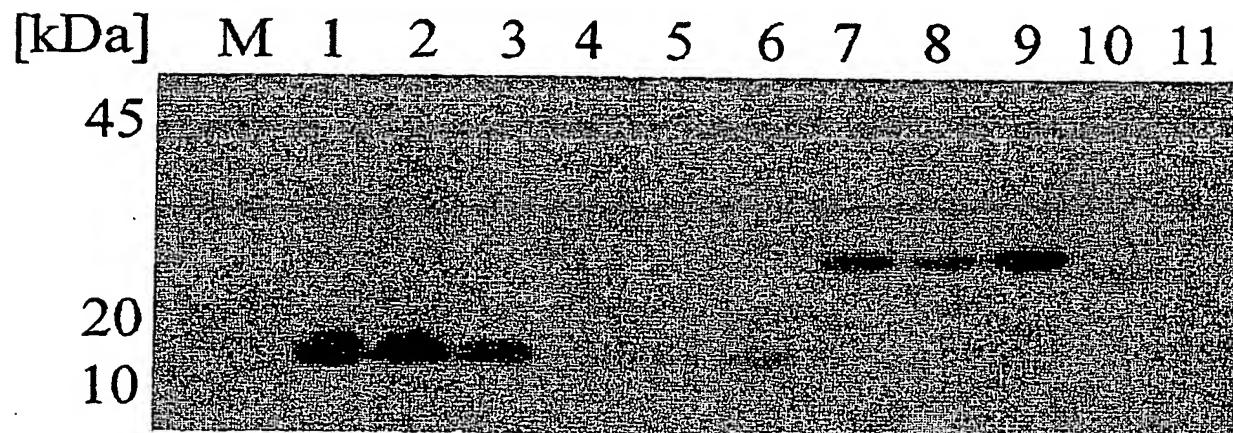
| Loop | Loop'   |
|------|---------|
| G A  | G A     |
| T T  | A T     |
| G-C  | T-A     |
| C-G  | A-T     |
| A-T  | A-T     |
| C-G  | A-T     |
| G-C  | C-G     |
| T-A  | A-T     |
| C-G  | G-C     |
|      | A-T     |
|      | C-G-T-A |

16/16

**Figur 13**

1049/Survivin/CIITA Loop und Loop'-Varianten

**Spur 1 1049 Mutante 1 Loop, Spur 2 Mutante 1 Loop', Spur 3 entspricht der 1049 Wildtypsequenz; Spur 4 Survivin Mutante 1 Loop, Spur 5 Mutante 1 Loop', Spur 6 entspricht der Survivin Wildtypsequenz; Spur 7 CIITA Mutante 1 Loop, Spur 8 Mutante 1 Loop', Spur 9 entspricht der CIITA Wildtypsequenz**

**Figur 14**

## SEQUENCE LISTING

<110> Roche Diagnostics GmbH

<120> Optimierte Proteinsynthese

<130> 29415pwo

<140> PCT/EP03

<141> 2003-12-09

<160> 57

<170> PatentIn Ver. 2.1

<210> 1

<211> 84

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Primer C

<400> 1

gaaatataata cgactcacta tagggagacc acaacggttt ccctctagaa ataattttgt 60  
ttaacttaa gaaggagata tacc 84

<210> 2

<211> 71

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Primer D

<400> 2

caaaaaaaccctcaagaccc gtttagagggc cccaaaggggg gcccgcaggta tgctgaattc 60  
gccttttatt a 71

<210> 3

<211> 30

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Primer A

without hairpinloop

<400> 3

aggagatata ccatgactag caaaggagaa

30

<210> 4

<211> 42

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Primer A  
Stammlänge 4 bp

<400> 4

aggagatata ccatgactaa ttttagtact agcaaaggag aa

42

<210> 5

<211> 45

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Primer A  
Stammlänge 5 bp

<400> 5

aggagatata ccatgactgt ttatacagta actagcaaag gagaa

45

<210> 6

<211> 48

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Primer A  
Stammlänge 6 bp

<400> 6

aggagatata ccatgactgg tcaattacca gtaactagca aaggagaa

48

<210> 7

<211> 51

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Primer A  
Stammlänge 7 bp

<400> 7

aggagatata ccatgactgc tttacatcaa gcagtaacta gcaaaggaga a

51

<210> 8

<211> 51

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Primer A  
Stammlänge 8 bp

<400> 8

aggagatata ccatgactgc acgtgatcgt gcagtaacta gcaaaggaga a

51

<210> 9

<211> 30

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Primer B

<400> 9

attcgccctt tattaatgtat gatgtatgt

30

<210> 10

<211> 60

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Primer A

<400> 10

aggagatata ccatgacttag cactgcacgt gcatcgtgca gtgtaaaagg agaagaactt 60

<210> 11

<211> 63

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Primer A

<400> 11

aggagatata ccatgactag caaaaactgca cgtgcacatcggt gcagtgttagg agaagaactt 60  
ttc 63

<210> 12

<211> 66

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Primer A

<400> 12

aggagatata ccatgactag caaaggaact gcacgtgcatt cgtgcagtgt gcgtt 60  
ttcaact 66

<210> 13

<211> 69

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Primer A

<400> 13

aggagatata ccatgactag caaaggagaa actgcacgtg catcgatcgatgtt 60  
ttcaactgg 69

<210> 14

<211> 72

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Primer A

<400> 14

aggagatata ccatgactag caaaggagaa gaaactgcac gtgcacatcggtt cagtgtactt 60

ttcaactggag tt

72

<210> 15  
<211> 75  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:Primer A

<400> 15  
aggagatata ccatgactag caaaggagaa gaacttactg cacgtgcac 60  
ttcaactggag ttgtc 75

<210> 16  
<211> 71  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:Primer D

<400> 16  
caaaaaaccc ctcaagaccc gtttagaggc cccaaagggt tgggagtaga atgttaagga 60  
ttagtttatt a 71

<210> 17  
<211> 60  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:Primer A

<400> 17  
aggagatata ccatgaaata tacatattct ctgcacgtga tcgtgcaggc taacaccgcg 60

<210> 18  
<211> 60  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:Primer A

<400> 18

aggagatata ccatgaaaac atattattct ctgcacgtga tcgtgcaggc taacaccg 60

<210> 19

<211> 60

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Primer A

<400> 19

aggagatata ccatgaaata ttcttataca ctgcacgtga tcgtgcaggc taacaccg 60

<210> 20

<211> 60

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Primer A

<400> 20

aggagatata ccatgaaata ttattctaca ctgcacgtga tcgtgcaggc taacaccg 60

<210> 21

<211> 60

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Primer A

<400> 21

aggagatata ccatgaaata tacatattca ctgcacgtga tcgtgcaggc taacaccg 60

<210> 22

<211> 60

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Primer A

<400> 22

aggagatata ccatgaaaac atattattca ctgcacgtga tcgtgcaggg taacaccg 60

<210> 23

<211> 60

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Primer A

<400> 23

aggagatata ccatgaaata ttcatataca ctgcacgtga tcgtgcaggg taacaccg 60

<210> 24

<211> 60

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Primer A

<400> 24

aggagatata ccatgaaata ttattcaaca ctgcacgtga tcgtgcaggg taacaccg 60

<210> 25

<211> 60

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Primer A

<400> 25

aggagatata ccatgcatac tcatacatcat ctgcacgtga tcgtgcaggg taacaccg 60

<210> 26

<211> 27

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Primer Wildtyp

<400> 26

aggagatata ccatggctaa caccgcg

27

<210> 27

<211> 48

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Primer B

<400> 27

aggatttagtt tattaatgat gatgatgatg atggcgccgg gtgcgcga

48

<210> 28

<211> 60

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Primer A

<400> 28

aggagatata ccatgaaata tacatattct ctgcacgtga tcgtgcaggg tgcccccacg 60

<210> 29

<211> 60

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Primer A

<400> 29

aggagatata ccatgaaaac atattattct ctgcacgtga tcgtgcaggg tgcccccacg 60

<210> 30

<211> 60

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Primer A

<400> 30  
aggagatata ccatgaaata ttcttataca ctgcacgtga tcgtgcaggg tgcccccacg 60

<210> 31  
<211> 60  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:Primer A

<400> 31  
aggagatata ccatgaaata ttattctaca ctgcacgtga tcgtgcaggg tgcccccacg 60

<210> 32  
<211> 60  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:Primer A

<400> 32  
aggagatata ccatgaaata tacatattca ctgcacgtga tcgtgcaggg tgcccccacg 60

<210> 33  
<211> 60  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:Primer A

<400> 33  
aggagatata ccatgaaaac atattattca ctgcacgtga tcgtgcaggg tgcccccacg 60

<210> 34  
<211> 60  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:Primer A

<400> 34  
aggagatata ccatgaaata ttcatataca ctgcacgtga tcgtgcaggg tgcccccacg 60

<210> 35  
<211> 60  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:Primer A

<400> 35  
aggagatata ccatgaaata ttattcaaca ctgcacgtga tcgtgcaggg tgcccccacg 60

<210> 36  
<211> 60  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:Primer A

<400> 36  
aggagatata ccatgcatca tcattcatcat ctgcacgtga tcgtgcaggg tgcccccacg 60

<210> 37  
<211> 27  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:Primer A  
Wildtyp

<400> 37  
aggagatata ccatgggtgc cccgacg

27

<210> 38  
<211> 49  
<212> DNA  
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Primer B

<400> 38

aggatttagtt tattaatgat gatgatgatg atgatccatg gcagccagc

49

<210> 39

<211> 60

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Primer

<400> 39

aggagatata ccatgaaata tacatattct ctgcacgtga tcgtgcagga gttggggccc 60

<210> 40

<211> 60

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Primer

<400> 40

aggagatata ccatgaaaac atattattct ctgcacgtga tcgtgcagga gttggggccc 60

<210> 41

<211> 60

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Primer

<400> 41

aggagatata ccatgaaata ttcttataca ctgcacgtga tcgtgcagga gttggggccc 60

<210> 42

<211> 60

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Primer

<400> 42

aggagatata ccatgaaata ttattctaca ctgcacgtga tcgtgcagga gttggggccc 60

<210> 43

<211> 60

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Primer

<400> 43

aggagatata ccatgaaata tacatattca ctgcacgtga tcgtgcagga gttggggccc 60

<210> 44

<211> 60

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Primer

<400> 44

aggagatata ccatgaaaac atattattca ctgcacgtga tcgtgcagga gttggggccc 60

<210> 45

<211> 60

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Primer

<400> 45

aggagatata ccatgaaata ttcatataca ctgcacgtga tcgtgcagga gttggggccc 60

<210> 46

<211> 60

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Primer

<400> 46

aggagatata ccatgaaata ttattcaaca ctgcacgtga tcgtgcagga gttggggccc 60

<210> 47

<211> 60

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Primer

<400> 47

aggagatata ccatgcata tcatacatcat ctgcacgtga tcgtgcagga gttggggccc 60

<210> 48

<211> 27

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Primer A

Wildtyp

<400> 48

aggagatata ccatggagtt ggggccc

27

<210> 49

<211> 45

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Primer B

<400> 49

aggatttagtt tattatataat gatgatgatg atgatgagaa cccccc

45

<210> 50

<211> 431

<212> DNA

<213> Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence:  
Expressionskonstrukt für Mutante 1

&lt;400&gt; 50

gaaattaata cgactcacta tagggagacc aacaacggttt ccctctagaa ataattttgt 60  
ttaactttaa gaaggagata taccatgaaa tatacatatt ctctgcacgt gatcgtgcag 120  
gctaacacccg cgccgggacc cacgggtggcc aacaaggcggg acgaaaaaca ccgtcacgtc 180  
gttaacgtcg ttttggagct gccgaccgag atatcagagg ccaccaccc ggtgttggcc 240  
accatgctga gcaagtacac gcgcatgtcc agcctgtta atgacaagtg cgcctttaag 300  
ctggacctgt tgcgcatggt agccgtgtcg cgacccggc gccatcatca tcatacatcat 360  
taataaacta atccttaaca ttctactccc aacccttgg ggcctctaaa cgggtcttga 420  
ggggttttttt g 431

&lt;210&gt; 51

&lt;211&gt; 398

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence:  
Expressionskonstrukt für Wildtyp

&lt;400&gt; 51

gaaattaata cgactcacta tagggagacc aacaacggttt ccctctagaa ataattttgt 60  
ttaactttaa gaaggagata taccatggct aacaccgcgc cgggacccac ggtggccaac 120  
aaggcgggacg aaaaacaccg tcacgtcggt aacgtcggtt tggagctgcc gaccgagata 180  
tcagaggcca cccacccggc gttggccacc atgctgagca agtacacgcg catgtccagc 240  
ctgtttaatg acaagtgcgc cttaagctg gacctgttgcc gcatggtagc cgtgtcgccg 300  
acccggcgcc atcatcatca tcatacattaa taaactaatac cttaacattc tactccaaac 360  
cccttggggc ctctaaacgg gtcttgaggg gttttttt 398

&lt;210&gt; 52

&lt;211&gt; 632

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence:  
Expresionskonstrukt Mutante 1

&lt;400&gt; 52

gaaattaata cgactcacta tagggagacc aacaacggttt ccctctagaa ataattttgt 60  
ttaactttaa gaaggagata taccatgaaa tatacatatt ctctgcacgt gatcgtgcag 120  
ggtccccca cggtcccccc tgcctggcag cccttctca aggaccaccg catctctaca 180  
ttcaagaact ggccttctt ggagggctgc gcctgcaccc cggagcggat ggccgaggct 240

ggcttcatcc actgccccac tgagaacgag ccagacttgg cccagtgtt cttctgcttc 300  
aaggagctgg aaggctgggac gccagatgac gacccatag aggaacataa aaagcattcg 360  
tccgggtcg cttccttgc tgtcaagaag cagtttgaag aattaaccct tggtaattt 420  
ttgaaactgg acagagaaaag agccaagaac aaaattgcaa aggaaaccaa caataagaag 480  
aaagaatttgc agggaaactgc gaagaaagtgc cgccgtgcac tcgagcagct ggctgccc 540  
gatcatcatc atcatcatca ttaataaaact aatccttaac attctactcc caacccttg 600  
gggcctctaa acgggtcttg aggggtttt tg 632

<210> 53  
<211> 599  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
Expressionskonstrukt Wildtyp

<400> 53  
gaaattaata cgactcacta tagggagacc acaacgggtt ccctctagaa ataattttgt 60  
ttaactttaa gaaggagata taccatgggt gccccgacgt tgccccctgc ctggcagccc 120  
tttctcaagg accaccgcatt ctctacattc aagaactggc ctttttggg gggctgcgc 180  
tgcacccccc agcggatggc cgaggctggc ttcatccact gccccactga gaacgagcca 240  
gacttggccc agtgtttctt ctgcttcaag gagcttggaaag gctgggagcc agatgacgac 300  
cccatagagg aacataaaaaa gcattcgtcc gtttgcgtt tcctttctgt caagaagcag 360  
tttgaagaat taacccttgg tgaatttttg aaactggaca gagaaagagc caagaacaaa 420  
attgcaaaagg aaaccaacaa taagaagaaa gaatttgagg aaactgcgaa gaaagtgcgc 480  
cgtgccatcg agcagctggc tgccatggat catcatcatc atcatcatta ataaactaat 540  
ccttaacatt ctactccaa ccccttgggg cctctaaacg ggttttgagg gtttttttg 599

<210> 54  
<211> 1400  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
Expressionskonstrukt Mutante 1

<400> 54  
gaaattaata cgactcacta tagggagacc acaacgggtt ccctctagaa ataattttgt 60  
ttaactttaa gaaggagata taccatggaaa tatacatatt ctctgcacgt gatcgtgcag 120  
gagttggggc ccctagaagg tggctacctg gagtttcta acagcgatgc tgacccctg 180  
tgcctctacc acttctatga ccagatggac ctggctggag aagaagagat tgagctctac 240  
tcagaacccg acacagacac catcaactgc gaccagtca gcaggctgtt gtgtgacatg 300  
gaaggtgatg aagagaccag ggaggcttat gccaatatcg cggaactggc ccagtatgtc 360  
ttccaggact cccagctggc gggcctgagc aaggacattt tcaagcacat aggaccagat 420

gaagtgtatcg gtgagagatggagatgc caaagtttgc ggcagaaaag tcagaaaaaga 480  
 ccctcccaag aggagcttcc ggcagacactg aagcaactgaa agccagctga gccccccact 540  
 gtggtactg gcagtctcct agtgggacca gtgagcgact gctccacccct gcccctgcctg 600  
 ccactgcctg cgctgttcaa ccaggagcca qcctccggcc agatgcgcct ggagaaaacc 660  
 gaccagattc ccatgcctt ctccagttcc tcgttgagct gcctgaatct ccctgaggg 720  
 cccatccagt ttgtccccac catctccact ctgccccatg ggctctggca aatctctgag 780  
 gctggaacag gggctccag tatattcatac taccatgggt aggtgcggca ggccagccaa 840  
 gtacccctc ccagtggatt cactgtccac ggcctccaa catctccaga cggccaggc 900  
 tccaccagcc ccttcgtctc atcagccact gacctgccc gcatgcctga acctgcctg 960  
 acctcccgag caaacatgac agagcacaag acgtccccca cccaatgcggc ggcagctgga 1020  
 gaggtctcca acaagcttcc aaaatggcct gagccgggtgg agcagttcta ccgctcactg 1080  
 caggacacgt atggtgcga gcccgcaggc ccggatggca tcctagtgga ggtggatctg 1140  
 gtgcaggcca ggctggagag gagcagcagc aagagcctgg agcgggaact ggccaccccg 1200  
 gactggcag aacggcagct ggcccaagga ggccctggctg aggtgctgtt ggctgccaag 1260  
 gagcaccggc ggccgcgtcg actcgagcga gctccgggg ggggttctca tcatacatcat 1320  
 catcatataat aataaactaa tccttaacat tctactccca acccccttggg gcctctaaac 1380  
 gggtctttag gggtttttg 1400

&lt;210&gt; 55

&lt;211&gt; 1367

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence:  
 Expressionskonstrukt Wildtyp

&lt;400&gt; 55

gaaattaata cgactcacta tagggagacc acaacggttt ccctctagaa ataattttgt 60  
 ttaactttaa gaaggagata taccatggag ttggggccccc tagaaggtgg ctacctggag 120  
 cttcttaaca gcgatgctga cccctgtgc ctctaccact tctatgacca gatggacctg 180  
 gctggagaag aagagattga gctctactca gaacccgaca cagacaccat caactgcgcac 240  
 cagttcagca ggctgttgg tgacatggaa ggtgatgaag agaccaggga ggcttatgcc 300  
 aatatcgccg aactggacca gtatgtttc caggactccc agctggaggg cctgagcaag 360  
 gacattttca agcacatagg accagatgaa gtgatcggtg agagtatgga gatgccagca 420  
 gaagttgggc agaaaagtca gaaaagaccc ttcccagagg agcttccggc agacctgaag 480  
 cactggaagc cagctgagcc ccccaactgtg gtgactggca gtctcttagt gggaccagt 540  
 agcgactgct ccaccctgcc ctgcctgcca ctgcctgcgc tggtaaccca ggagccagcc 600  
 tccggccaga tgcgcctgga gaaaaccgac cagattccca tgccttctc cagttcctcg 660  
 ttgagctgcc tgaatctccc tgagggaccc atccagtttgc tccccaccat ctccactctg 720  
 ccccatgggc tctggcaat ctctgaggct ggaacagggg tctccagtttattcatctac 780  
 catggtgagg tgccccaggc cagccaagta cccctccca gtggattcac tggccacggc 840  
 ctcccaacat ctccagaccg gccaggctcc accagccct tcgctccatc agccactgac 900  
 ctgcccagca tgcctgaacc tgcctgacc tcccggacaa acatgacaga gacacaagacg 960  
 tccccccaccc aatggccggc agctggagag gtctccaaca agcttccaaa atggcctgag 1020  
 ccgggtggagc agttctaccg ctcactgcag gacacgtatg gtgcccggcc cgcaggcccg 1080  
 gatggcatcc tagtggaggt ggtatcggtg caggccaggc tggagaggag cagcagcaag 1140

agcctggagc ggaaactggc caccggac tggcagaac ggcagctggc ccaaggaggc 1200  
ctggctgagg tgctgttggc tgccaggag caccggcgcc cgctcgact cgagcgagct 1260  
cccggggggg gtttcatca tcatacatcat catataataat aaactaatcc ttaacattct 1320  
actcccaacc ctttgggccc tctaaacggg tcttgagggg tttttt 1367

<210> 56  
<211> 938  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Expressionskonstrukt

<400> 56  
gaaattaata cgactcacta tagggagacc acaacggttt ccctctagaa ataattttgt 60  
ttaactttaa gaaggagata taccatgaaa tatacatatt ctctgcacgt gatcgtgcag 120  
actagcaaag gagaagaact tttcaactgga gttgtcccaa ttcttggta attagatgtt 180  
gatgttaatg ggcacaaatt ttctgtcagt ggagagggtg aaggtgatgc tacatacgg 240  
aagtttaccc ttaaattttt ttgcactact ggaaaactac ctgttccatg gccaacactt 300  
gtcactactt tctcttattgg tttcaatgc ttttcccggtt atccggatca tatgaaacgg 360  
catgactttt tcaagagtgc catgcccggaa gtttatgtac aggaacgcac tatatcttc 420  
aaagatgacg ggaactacaa gacgcgtgct gaagtcaagt ttgaaggtga tacccttgg 480  
aatcgatcg agttaaaaagg tattgatttt aaagaagatg gaaacattct cggacacaaa 540  
ctcgagtaca actataactc acacaatgta tacatcacgg cagacaaaca aaagaatgga 600  
atcaaagcta acttcaaaat tcgcccacaac attgaagatg gatccgttca actagcagac 660  
cattatcaac aaaatactcc aattggcgat ggccctgtcc ttttaccaga caaccattac 720  
ctgtcgacac aatctgccc ttcgaaagat cccaacgaaa agagagacca catggccctt 780  
cttgagtttg taacagctgc tggattaca catggcatgg atgaactata caaaccggg 840  
gggggttctc atcatcatca tcatacataa taaactaatac cttaacattc tactccaaac 900  
cccttgggccc ctctaaacgg gtcttgaggg gttttt 938

<210> 57  
<211> 905  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
Expressionskonstrukt

<400> 57  
gaaattaata cgactcacta tagggagacc acaacggttt ccctctagaa ataattttgt 60  
ttaactttaa gaaggagata taccatgact agcaaaggag aagaactttt cactggagg 120  
gtcccaattt ttgttgaattt agatgggtat gttaatgggc acaaattttc tgcgtgg 180  
gagggtgaag gtgtatgtac atacggaaag ctaccctta aatttatttg cactactgga 240

18/18

aaactacctg ttccatggcc aacacttgc actactttct cttatggtgt tcaatgcctt 300  
tcccgttatac cgatcatat gaaacggcat gacttttca agagtccat gcccgaaggt 360  
tatgtacagg aacgcactat atcttcaaa gatgacggga actacaagac gcgtgctgaa 420  
gtcaagtttgc aaggtgatac cttgttaat cgtatcgagt taaaaggat tgattttaaa 480  
gaagatggaa acattctcg acacaaactc gagtacaact ataactcaca caatgtatac 540  
atcacggcag acaaacaaaa gaatgaaatc aaagctaact tcaaaaattcg ccacaacatt 600  
gaagatggat ccgttcaact agcagaccat tatcaacaaa atactccat tggcgatggc 660  
cctgtcctt taccagacaa ccattacctg tcgacacaat ctgccttgc gaaagatccc 720  
aacgaaaaga gagaccacat ggtccttctt gagtttgtaa cagctgctgg gattacacat 780  
ggcatggatg aactatacaa acccgggggg ggttctcatc atcatcatca tcattaataa 840  
actaattcctt aacattctac tcccaacccc ttggggcctc taaacgggtc ttgaggggtt 900  
ttttg 905